A new Technique for Calibrating Stochastic Volatility Models: The Malliavin Gradient Method

Ewald, Christian-Oliver; Zhang, Aihua

QUANTITATIVE FINANCE

A new Technique for Calibrating Stochastic Volatility Models : The Malliavin Gradient Method

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2005-0015.R1
Manuscript Category:	Research Paper
Date Submitted by the Author:	26-Oct-2005
Complete List of Authors:	Ewald, Christian-Oliver; University of Leeds, School of Mathematics Zhang, Aihua; University of Leeds, Leeds University Business School
Keywords:	Monte Carlo Methods, Calibration of Stochastic Volatility, Derivative Pricing Models, Option Pricing via Simulation, Computational Finance, Stochastic Volatility, Financial Mathematics, Value at Risk
JEL Code:	C00 - General < C0 - General < C - Mathematical and Quantitative Methods, C61 - Optimization Techniques|Programming Models|Dynamic Analysis < C6 - Mathematical Methods and Programming < C - Mathematical and Quantitative Methods, C51 - Model Construction and Estimation < C5 - Econometric Modeling < C - Mathematical and Quantitative Methods, C19 - Other < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, C15 - Statistical Simulation Methods|Monte Carlo Methods < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods

A new Technique for Calibrating Stochastic Volatility Models : The Malliavin Gradient Method

Christian-Oliver Ewald
School of Mathematics
University of Leeds
Leeds, LS2 9JT, UK

Aihua Zhang
Leeds University Business School
University of Leeds
Leeds, LS2 9JT, UK

revised, October 2005

Abstract

We discuss the application of gradient methods to calibrate mean reverting stochastic volatility models. For this we use formulas based on Girsanov transformations as well as a modification of the Bismut-Elworthy formula to compute the derivatives of certain option prices with respect to the parameters of the model by applying Monte Carlo methods. The article presents an extension of the ideas to apply Malliavin calculus methods in the computation of Greek's.

Keywords: Malliavin calculus, Monte Carlo simulation, stochastic volatility models, calibration, gradient methods, value at risk

JEL Subject Classification: C00, C15, C19, C51, C61

Introduction

Models for financial markets depend in general on certain parameters, which one can choose in a way that the market model resembles the corresponding real market best. A simple one dimensional standard Black-Scholes model would depend for example on the drift term b of the stock as well as the volatility coefficient σ and a natural question would be : How should one choose b and σ such that the option prices predicted by this Black-Scholes model are closest to observed market prices ? As is well known, under risk neutral valuation b does not affect prices and so the minimization affects only the parameter σ alone. For most elementary options this problem could be solved explicitly, however for certain exotic options it would be hard enough to solve, at least when one is trying to get closed form solutions. Many people working in Mathematical Finance do not consider the Black-Scholes model as an appropriate market model anymore. Evidence for this is given for example by volatility smiles. Nowadays many financial institutions work with so called mean reverting stochastic volatility models. As the simple Black-Scholes model involves certain parameters, so do mean reverting stochastic volatility models and the question remains : How to chose the parameters in a way that the option prices predicted by the models best fit observed market prices ? Solving this problem by gradient methods from optimization is the main theme of this article. We allow our models to depend on the four parameters κ, θ, γ_1 and γ_2. The parameters κ and θ are classic parameters describing the mean reversion, γ_1 is a measure for the correlation between stock and volatility and γ_2 is a parameter describing the volatility of the volatility. When applying gradient methods we have to differentiate the option prices $\Pi(\kappa, \theta, \gamma)(h(S_T))$ of a certain European style option on a stock (S_t) with payoff function h with respect to κ, θ and γ. Problems occur when one has to differentiate h, since almost all options possess payoff functions with singularities. In our approach we avoid differentiating h by using a certain combination of Girsanov drift adjustment and Malliavin's integration by parts formula. These ideas build up on ideas by Fournie et al. used for the computation of Greeks (see [7]). We give formulas for the gradient of $\Pi(\kappa, \theta, \gamma)(h(S_T))$ which do not involve differentiating h and are accessible by Monte Carlo methods. For the simulations we use plain Monte Carlo simulation, which works well in our examples. The numerical results can be improved by using the general machinery of importance sampling or stratified sampling as well as certain variance reduction techniques. Since these are standard techniques, well known to practical analysts and financial programmers, we leave this aspect out of our considerations. The main goal of the article is to illustrate the new idea, how to use Malliavin calculus and

standard gradient methods for calibration. The presentation is given in a way, that it is easy work for the practitioner to adjust the method and to optimize it for his or her special needs.

The article is organized as follows. In section 1 we present the necessary background about mean reverting stochastic volatility models as well as option pricing in such models. We consider the general problem of calibration of such models in section 2. Using techniques from Malliavin calculus we prove a modification of the Bismut-Elworthy formula in section 3. Section 4 provides the formulas for the partial derivatives of the option prices $\Pi(\kappa, \theta, \gamma)(h(S_T))$ with respect to the parameters κ, θ and γ and in section 5 we implement the combined Malliavin gradient algorithm. Section 7 summarizes the main ideas and conclusions of the article. As an example how the Malliavin gradient method can be implemented, we study the case of a Vasicek model in Appendix A. We give explicit formulas for this case and present some simulations. Appendices B and C give background information on gradient methods resp. Malliavin calculus, only as far as is needed to read this article.

All simulations have been done by using the *Matlab*© software package. Our thanks go to the Stochastic Control and Mathematical Finance Group of the University of Kaiserslautern for presenting us with a rich and stimulating working environment and Ralf Korn, Heinrich von Weizsäcker,David Nualart and Klaus Schenk-Hoppé for fruitful discussions and advice. Additionally the first author wants to thank in particular the BBSRC and the University of Bristol for financial support. The second author wants to thank Leeds University Business School as well as the University of Kaiserslautern for financial support. Thanks also go to an anonymous referee for many useful comments and suggestions which we used to improve an earlier version of this article.

1 Option pricing in mean reverting stochastic volatility models

In this section we introduce our market model. We assume right from the beginning, that we work under a risk neutral measure. As is well known, this is the right framework when doing option pricing. We concentrate on so called mean reverting stochastic volatility models, which are widely used among practitioners in finance. Such models are characterized by an affine drift term of the form $\kappa(\theta - v_t)dt$ in the SDE for the volatility. A detailed

discussion of such models can be found for example in [8]. Within the whole article we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is rich enough to carry a Brownian motion of at least dimension two and is filtered by the corresponding Wiener filtration, which we denote with $(\mathcal{F}_t)$. We can always think of a multidimensional Wiener space.

The models we consider are simple in a way that they only involve one stock and the volatility process. Stock and volatility are assumed to follow the dynamics

$$\begin{aligned} dS_t &= S_t(rdt + f(v_t)dW_t) \\ dv_t &= \kappa(\theta - v_t)dt + g(v_t, \nu)d\tilde{W} \end{aligned}$$

where for simplicity we assume that the model parameters κ,θ and ν are positive constants. The interest rate r is assumed to be deterministic and by considering discounted stock prices instead of actual stock prices we can as well assume that $r \equiv 0$, which we will do from now on. The functions f and g are supposed to be smooth with bounded derivatives of all orders. These are standard assumptions when Malliavin calculus methods are applied to certain diffusion processes. A typical example where these assumptions apply is the Vasicek model (see [8], page 177). In some situations these assumptions can be relaxed using localization or drift adjustment techniques as in [5] and [3]. This is the case for example in the Heston model [13] and the Hull and White model [14]. The two Brownian motions W and $\tilde{W}$ are allowed to be correlated in a way that there exists a constant $-1 < \rho < 1$ such that

$$d < W_t, \tilde{W}_t >= \rho dt.$$

Setting $\mu := \sqrt{1-\rho^2}$ one can find a two dimensional Brownian motion $\mathbb{W} = (W^1, W^2)^\top$ such that the dynamics of stock and volatility can be written as

$$\begin{aligned} dS_t &= S_t f(v_t)\left(\mu dW_t^1 + \rho dW_t^2\right) \\ dv_t &= \kappa(\theta - v_t)dt + g(v_t, \nu)dW^2. \end{aligned}$$

Setting $X_t = \begin{pmatrix} S_t \\ v_t \end{pmatrix}$ and using a two dimensional notation we can write

$$dX_t = \beta(X_t)dt + \sigma(X_t)d\mathbb{W}_t$$

where $\beta(x_1, x_2) = \begin{pmatrix} 0 \\ \kappa(\theta - x_2) \end{pmatrix}$ and $\sigma(x_1, x_2) = \begin{pmatrix} \mu x_1 f(x_2) & \rho x_1 f(x_2) \\ 0 & g(x_2, \nu) \end{pmatrix}$.

In our setup, we assume that agents in our model are aware of the general mean reverting structure of the model, including the functions f and g, but not on the precise numerical values of the model parameters κ, θ, ν and ρ. This feature leads to the problem of calibration, which we are going to discuss in the next section. However to include the dependence of the model on these parameters we include them in the notation of the functions β and σ, i.e. we write $\beta(x_1, x_2, \kappa, \theta)$ as well as $\sigma(x_1, x_2, \rho, \nu)$. We also set $\gamma = (\rho, \nu) \in \mathbb{R}^2$. Certainly the process (X_t) depends on the parameters κ, θ and γ and we think of X_t as a family of stochastic processes parametrized by κ, θ, γ defining solutions to the family of stochastic differential equations

$$dX_t = \beta(X_t, \kappa, \theta)dt + \sigma(X_t, \gamma)d\mathbb{W}_t$$

(see [19], Theorem 39 for the existence of families of solutions). In addition to the parameters introduced one can add the initial condition as another parameter. The dependence of option prices on initial conditions however, already has been studied extensively. In this article we fix initial conditions as

$$x = \begin{pmatrix} S_0(\kappa, \theta, \gamma) \\ v_0(\kappa, \theta, \gamma) \end{pmatrix} = X_0(\kappa, \theta, \gamma)$$

for all κ, θ, γ. Instead of a family of stochastic differential equations, one can as well think of a family of market models described by the parameters κ, θ, γ and a natural question is, which of the models within the family one should chose.

Let us now consider the problem of pricing a contingent claims of the form $h(S_T)$ where $h : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and T is the expiry time. Such contingent claims correspond to European style options. Since we chose a risk neutral setup in the beginning, we can compute an arbitrage free price at time $t = 0$ via the following formula

$$\Pi(\kappa, \theta, \gamma)(h(S_T)) := \mathbb{E}_{\mathbb{P}}(h(S_T(\kappa, \theta, \gamma)))$$

This price naturally depends on the parameters κ, θ and γ. The parameters κ and θ determining the drift term of the volatility and in a way represent the agents choice of the risk neutral measure, as long as a mean reverting structure is assumed. The parameter γ reflects the agents assessment of the correlation and volatility of volatility.

2 Calibration of stochastic volatility models

Options on stocks and other financial derivatives are traded on financial markets throughout the world. Often traders use certain models to price derivatives, but not always consistently and even worse, they do not tell us which model they use. A good model should reflect the market prices. There are conceptual problems with the widely used Black-Scholes model and it is common knowledge that mean reverting stochastic volatility models solve some of these problems. The question however remains how to determine the parameters of the model in a way that the approximation to market prices is best. In our setting this means we have to determine the parameters κ, θ and γ in a way that our choice minimizes a certain functional which measures the distance of model prices to market prices. This is a classical optimization problem. From an economic point of view the choice of the parameters κ and θ is of particular interest, since this choice correspond to the martingale measure chosen by the traders within the market. A lot of theoretical research has been done to characterize equivalent martingale measures by certain criteria. Only to mention a few, there are minimal distance martingale measure (see [11],[12]), minimal entropy martingale measures (see [9]) or minimax martingale measures (see [1], [11]). These considerations however, do not help to determine which martingale measures are actually used on the market. Our considerations do not only affect the choice of an equivalent martingale measure, but also choices of parameters determining the volatility. Let us assume, that we have a mean reverting stochastic volatility model described by the parameters κ, θ and γ which we would like to test on real market data. Then we can consider a number n of European style options $h_i(S_{T_i})$ with maturity times T_i and their observed market-prices $\Pi^{obs}(h_i(S_{T_i}))$. The functional

$$L(\kappa, \theta, \gamma) = \sum_{i=1}^{n} \left[\Pi(\kappa, \theta, \gamma)(h_i(S_{T_i})) - \Pi^{obs}(h_i(S_{T_i}))\right]^2$$

is a measure for the error corresponding to the model described by the parameters κ, θ and γ. Calibration of our model now means to find minimizers for this functional, i.e.

$$(\kappa^*, \theta^*, \gamma^*) = \text{ argmin } L(\kappa, \theta, \gamma).$$

In section 5 we will demonstrate a method for the computation of κ^*, θ^* and γ^*.

3 First variation processes and a modified Bismut-Elworthy formula

In this section we apply Malliavin calculus techniques in order to derive a modified version of the Bismut-Elworthy formula. The construction of the Malliavin derivative operator as well as its fundamental properties are summarized in Appendix C. Let us consider the stochastic differential equation

$$dX_t = \beta(X_t)dt + \sigma(X_t, \gamma)d\mathbb{W}_t.$$

where $(\mathbb{W}_t)$ is an n-dimensional Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\gamma \in \mathbb{R}^n$ is a parameter determining the volatility of the solutions. We assume that the coefficient functions $\beta : \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times \mathbb{R}^n \to GL(n, \mathbb{R})$ are two times continuously differentiable and denote with $D^x\beta$ the Jacobian of β with respect to x as well as with $D^x\sigma_j$ respectively $D^\gamma\sigma_j$ the Jacobians of the j-th column of σ with respect to x respectively γ. The function β could be allowed to depend on more parameters, as in the previous sections where it depended on κ and θ, but this is not of importance for our considerations in this section, which will deal exclusively with the parameter γ. We therefore omit any possible dependencies of β on additional parameters from the notation. However, under these assumptions it follows from Theorem 39, page 305 in [19] that there exists a family of solutions $(X_t(x, \gamma))$ such that $X_0(x, \gamma) = x$ a.s. and the map

$$(x, \gamma) \mapsto X_t(x, \gamma)(\omega)$$

is continuously differentiable a.s. for all t defined up to an explosion time $\zeta(x, \gamma)$. We assume here that this explosion time is infinity for all choice of x and γ and moreover that the process (X_t) belongs to $L^2(\Omega \times [0, T], \mathbb{R}^n)$. Criteria to assure this can be found in [15] (Theorem 2.4 Chapter IV, page 163). Then the $\mathbb{R}^{n\times n}$ valued processes (D^xX_t) and $(D^\gamma X_t)$ satisfy

$$dD^xX_t = D^x\beta(X_t)D^xX_t dt + \sum_{j=1}^{n} D^x\sigma_j(X_t, \gamma)D^xX_t dW_t^j$$

with $D^xX_0 = Id$ a.s. as well as

$$dD^\gamma X_t = D^x\beta(X_t)D^\gamma X_t dt + \sum_{j=1}^{n} D^x\sigma_j(X_t, \gamma)D^\gamma X_t dW_t^j + \sum_{j=1}^{n} D^\gamma\sigma_j(X_t, \gamma)dW_t^j$$

with $D^\gamma X_0 = 0$. Here 0 denotes the $n \times n$ zero matrix whereas above Id denotes the $n \times n$ identity matrix. The process (D^xX_t) is often called

the first variation process corresponding to (X_t) and is in fact a $GL(n,\mathbb{R})$-valued process. The process $(D^\gamma X_t)$ so far has not been appeared in the literature and therefore has not been named. We refer to this process as γ-variation process. Both, the stochastic differential equation for the first variation process as well as the stochastic differential equation for the γ-variation process are linear and can therefore be solved explicitly in terms of the process (X_t). Using general results about the uniqueness of a solution to a linear stochastic differential equation, one also obtains that if one computes solutions to the stochastic differential equations above, then these solutions are necessarily the first variation respectively the γ-variation process. As it is easily verified the first variation process and the γ-variation process are related by the following formula :

$$\begin{aligned} D^\gamma X_t &= (D^x X_t)\cdot\left(\sum_{j=1}^n \int_0^t (D^x X_s)^{-1} D^\gamma \sigma_j(X_s,\gamma) dW_s^j \right. \\ &- \left. \sum_{j=1}^n \int_0^t (D^x X_s)^{-1} D^x \sigma_j(X_s,\gamma) D^\gamma \sigma_j(X_s,\gamma) ds \right) \end{aligned}$$

Furthermore, the first variation process is closely connected to the Malliavin derivative via the following proposition (compare [18], Theorem 2.2.1).

Proposition 3.1. *If under the assumptions from above $D^x X_t (D^x X_s)^{-1}\sigma(X_s,\gamma)$ is in $L^2(\Omega)$ for all $s\in[0,T]$, then $X_t \in \mathbb{D}_{1,2}$ and its Malliavin derivative is given by*

$$D_s X_t = D^x X_t (D^x X_s)^{-1}\sigma(X_s,\gamma)\cdot \mathbf{1}_{\{s\le t\}}$$

The relationship of the first variation process and the Malliavin derivative together with the integration by parts formula can be used to achieve the Bismut-Elworthy formula which appeared in several versions (see for example [18] page 125 or [7] Proposition 3.2).

Proposition 3.2. *If under the assumption of Proposition 3.1. the weight*

$$\pi_T = \frac{1}{T}\int_0^T \left(\sigma(X_t,\gamma)^{-1} D^x X_t\right)^\top d\mathbb{W}_t$$

is in $L^2(\Omega)$ and $h\in L^2_{loc}(\mathbb{R}^n,\mathbb{R})$ with at most linear growth at infinity, such that $h(X_T)\in L^2(\Omega)$, then

$$D^x\mathbb{E}(h(X_T)) = \mathbb{E}(h(X_T)\pi_T)$$

The Bismut Elworthy formula gives an expression for the derivative with respect to the initial condition and has so far very successfully been used to accelerate the computations of the Greek's Delta, Gamma and Vega (see [7]). These Greek's play a major role in hedging derivatives. The following proposition presents a modification of the Bismut-Elworthy formula concerning the derivative with respect to the parameter γ.

Proposition 3.3. *Under the assumptions of Proposition 3.1. define the weight*

$$\begin{aligned}\Gamma_T &:= \sum_{j=1}^{n}\int_0^T (D^x X_s)^{-1} D^\gamma \sigma_j(X_s,\gamma) dW_s^j \\ &- \sum_{j=1}^{n}\int_0^T (D^x X_s)^{-1} D^x \sigma_j(X_s,\gamma) D^\gamma \sigma_j(X_s,\gamma) ds\end{aligned}$$

and assume that the process $\tilde{\rho}_T$ defined by

$$\tilde{\rho}_{T,t} := \frac{1}{T}\sigma(X_t,\gamma)^{-1} D^x X_t \cdot \Gamma_T$$

belongs to dom(δ). Furthermore, let $h \in L^2_{loc}(\mathbb{R}^n,\mathbb{R})$ with at most linear growth at infinity such that $h(X_T) \in L^2(\Omega)$, then

$$D^\gamma \mathbb{E}(h(X_T)) = \mathbb{E}(h(X_T)\rho_T)$$

for the $\mathbb{R}^n$-valued weight $\rho_T := \delta(\tilde{\rho}_T)$.

Proof. Let us first assume the function h has bounded partial derivatives. Then one can interchange differentiation and expectation in order to obtain

$$D^\gamma \mathbb{E}(h(X_T)) = \mathbb{E}(\nabla h(X_T) D^\gamma X_T)$$

by the same argument as in the proof of Proposition 3.2. in [7]. By Proposition 3.1 we have

$$D_t X_T = D^x X_T (D^x X_t)^{-1} \sigma(X_t,\gamma)$$

or equivalently

$$D^x X_T = D_t X_T \sigma(X_t,\gamma)^{-1} D^x X_t$$

Integrating this equation over $t \in [0,T]$ and dividing by T gives

$$D^x X_T = \frac{1}{T}\int_0^T D_t X_T \sigma(X_t,\gamma)^{-1} D^x X_t dt.$$

The discussion preceding Proposition 3.1 showed $D^\gamma X_T = D^x X_T \cdot \Gamma_T$. Furthermore by definition $\tilde{\rho}_{T,t} = \frac{1}{T}\sigma(X_t,\gamma)^{-1} D^x X_t \cdot \Gamma_T$. Therefore, by multiplying the equation from above with Γ_T we obtain

$$D^\gamma X_T = D^x X_T \cdot \Gamma_T = \frac{1}{T}\int_0^T D_t X_T \sigma(X_t,\gamma)^{-1} D^x X_t \cdot \Gamma_T dt = \int_0^T (D_t X_T)\tilde{\rho}_{T,t} dt.$$

and furthermore

$$D^\gamma \mathbb{E}(h(X_T)) = \mathbb{E}\left(\nabla h(X_T)\int_0^T (D_t X_T)\tilde{\rho}_{T,t} dt\right).$$

By Proposition C3 and Proposition C2 in Appendix C we obtain

$$D^\gamma \mathbb{E}(h(X_T)) = \mathbb{E}\left(\int_0^T (D_t h(X_T))\tilde{\rho}_{T,t} dt\right) = \mathbb{E}(h(X_T)\delta(\tilde{\rho}_T))$$

and therefore $D^\gamma \mathbb{E}(h(X_T)) = \mathbb{E}(h(X_T)\rho_T)$.

Now consider the general case, when $h \in L^2_{loc}(\mathbb{R}^m,\mathbb{R})$ with at most linear growth at infinity. We prove the last formula by approximation of h by continuously differentiable functions as follows. First note that h can be written as $h = \tilde{h} + g$ where $\tilde{h} \in L^2(\mathbb{R}^n,\mathbb{R})$and g is continuously differentiable with bounded derivatives. By the argument above, our formula is correct for g and in order to prove our formula for h we only have to prove the formula for $\tilde{h}$ by additivity of the expectation. Therefore, we can as well assume that $h \in L^2(\mathbb{R}^n,\mathbb{R})$. Now since the algebra of smooth functions with compact support $C_c^\infty(\mathbb{R}^n,\mathbb{R})$ is dense in $L^2(\mathbb{R}^n,\mathbb{R})$, we can find a sequence of functions $h_n \in C_c^\infty(\mathbb{R}^n,\mathbb{R})$ such that $h_n \to h$ in $L^2(\mathbb{R}^n,\mathbb{R})$. It follows from Theorem 2.3.1, page 110 in [18], that for each triple (κ,θ,γ) the law of $X_T(\kappa,\theta,\gamma)$ is absolutely continuous with respect to the Lebesgue measure μ on $\mathbb{R}^n$. It therefore follows that $h_n(X_T(\kappa,\theta,\gamma)) \to h(X_T(\kappa,\theta,\gamma))$ in $L^2(\Omega)$ for each (κ,θ,γ). Using Egorovs Theorem ([23] page 16) it follows that the sequence h_n converges in fact μ-uniformly on compacts to h, where μ denotes the Lebesgue measure on $\mathbb{R}^n$. This can be used to show that the convergence

$$\mathbb{E}\left[(h_n(X_T) - h(X_T))^2\right] \to 0$$

is uniformly on compacts in the parameters κ,θ and γ. Denoting

$$\mathcal{C} := \mathbb{E}\left(h(X_T)\delta(\tilde{\rho}_T)\right)$$

we have by applying the obtained result for h_n as well as the Cauchy-Schwartz inequality that

$$\begin{aligned}\left|\frac{\partial}{\partial v}\mathbb{E}(h_n(X_T)) - \mathcal{C}\right| &= \left|\mathbb{E}\left((h_n(X_T) - h(X_T)) \cdot \delta(\tilde{\rho}_T)\right)\right| \\ &= \mathbb{E}\left[(h_n(X_T) - h(X_T))^2\right]^{1/2} \cdot \mathbb{E}\left[\delta(\tilde{\rho}_T)^2\right]^{1/2}\end{aligned}$$

As noted before, the first expression on the right hand side of the equality above converges uniformly (in κ, θ, γ) on compacts to zero. As a continuous function in (κ, θ, γ), the second expression on the right hand side of the equality is bounded on any compact set. It therefore follows that

$$D^\gamma \mathbb{E}(h_n(X_T)) \to \mathcal{C}$$

uniformly on compacts (in κ, θ, γ). A well-known result from basic calculus then implies that $\mathbb{E}(h(X_T))$ is differentiable with respect to γ and that the derivative is given by

$$D^\gamma \mathbb{E}(h(X_T)) = \mathcal{C} = \mathbb{E}\left(h(X_T)\delta(\tilde{\rho}_T)\right),$$

which was to prove.

□

The technical integrability conditions in Propositions 3.1, 3.2 and 3.3 are in general satisfied when the coefficient functions β and σ have bounded derivatives of all orders and σ is assumed to be uniformly coercitive. If this is not the case, one has to check these conditions for the individual case.

The explicit computation of the weight function $\rho_T := \delta(\tilde{\rho}_T)$ in the Proposition above can turn out to be difficult unless the Wiener chaos decomposition of the integrand is known. For approximation of Skorohod integrals in the one dimensional case see for example [18], page 150. To avoid Skorohod integrals at all, we use the following lemma.

Lemma 3.1. *Let e_l denote the l-th standard base vector in $\mathbb{R}^n$ and let (M_t) be an $\mathbb{R}^{n\times n}$-valued $(\mathcal{F}_t)$-adapted process such that for each component $(M_t)_{i,k}$ of (M_t) and each $1 \le l \le n$ the n-dimensional process $(M_t)_{i,k} \cdot e_l^\top$ is in $dom(\delta)$. Furthermore, let Γ be an $\mathcal{F}_T$ measurable random variable in $\mathbb{D}_{1,2}$ such that $\mathbb{E}\left(\|\Gamma\|^2 \int_0^T \|M_t\|^2 dt\right) < \infty$ where $\|\Gamma\|$ and $\|M_t\|$ denote the Euclidean norm*

of the matrices. Then the $\mathbb{R}^{n\times n}$-valued process $(M_t\cdot\Gamma)$ is Skorohod integrable and the i-th component of the $\mathbb{R}^n$-valued random variable $\delta(M_t\cdot\Gamma)$ is given by the following formula :

$$\delta(M_t\cdot\Gamma)_i=\sum_{l,k=1}^{n}\Gamma_{k,l}\int_0^T(M_t)_{i,k}dW_t^l-\sum_{l,k=1}^{n}\int_0^T D_t^l\Gamma_{k,l}\cdot(M_t)_{i,k}dt.$$

Here the integrals with respect to dW_s^j are ordinary Itô integrals.

Proof. The i-th component of $\delta(M_t\cdot\Gamma)$ is given by $\delta\left((M_t\cdot\Gamma)_{i,.}\right)$, where $(M_t\cdot\Gamma)_{i,.}$ denotes the $i-$th row of $M_t\cdot\Gamma$. We write this row as follows :

$$(M_t\cdot\Gamma)_{i,\cdot}=\sum_{l=1}^{n}\sum_{k=1}^{n}\Gamma_{k,l}\cdot(M_t)_{i,k}\cdot e_l^\top.$$

Using Proposition 4.9 in [20] (see also [18] formula 1.49, page 40) we obtain

$$\begin{aligned}\delta\left((M_t\cdot\Gamma)_{i,.}\right) &= \sum_{l=1}^{n}\sum_{k=1}^{n}\Gamma_{k,l}\delta\left((M_t)_{i,k}\cdot e_l^\top\right)\\ &\quad-\sum_{l=1}^{n}\sum_{k=1}^{n}<D\Gamma_{k,l}^\top,(M_t)_{i,k}\cdot e_l>_{L^2([0,T],\mathbb{R}^n)}\\ &= \sum_{l,k=1}^{n}\Gamma_{k,l}\int_0^T(M_t)_{i,k}dW_t^l-\sum_{l,k=1}^{n}\int_0^T D_t^l\Gamma_{k,l}\cdot(M_t)_{i,k}dt\end{aligned}$$

where in order to obtain the last equality we used Proposition C1 in Appendix C.

□

Using the previous lemma we obtain the following corollary of Proposition 3.2.

Corollary 3.1. *Assume that the assumptions in Lemma 3.1 hold for $\Gamma=\Gamma_T$ and (M_t) defined by*

$$(M_t)_{ik}=\frac{1}{T}\sum_{m=1}^{n}\sigma(X_t,\gamma)^{-1}_{im}\frac{\partial}{\partial x_k}X_t^m$$

for $i,k=1,..,n$. Then under the same assumptions as in Proposition 3.1 the i-th component of the $\mathbb{R}^n$ valued weight function ρ_T from Proposition 3.3 can be computed by the formula

$$\begin{aligned}\rho_{T,i} &= \frac{1}{T}\sum_{l,k,m=1}^{n}\Gamma_{T,k,l}\int_0^T \sigma(X_t,\gamma)^{-1}_{im}\frac{\partial}{\partial x_k}X_t^m dW_t^l \\ &- \frac{1}{T}\sum_{l,k,m=1}^{n}\left(D_t^l\Gamma_{T,k,l}\right)\sigma(X_t,\gamma)^{-1}_{im}\frac{\partial}{\partial x_k}X_t^m dt.\end{aligned}$$

Proof. This follows immediately from Proposition 3.2 and Lemma 3.1. □

4 The derivatives of option prices with respect to the model parameters

Let us now reconsider the family of mean reverting stochastic volatility models discussed in section 1:

$$dX_t = \beta(X_t,\kappa,\theta)dt + \sigma(X_t,\gamma)d\mathbb{W}_t,$$

In this section, we study the derivatives of the option price $\Pi(\kappa,\theta,\gamma)(h(S_T))$ with respect to the parameters κ,θ,γ. These derivatives will later be used to apply gradient methods for the search of the "best" model within the family. As in section 3, it follows from our assumptions and Theorem 39 in [19] that we can assume w.l.o.g. that for each t and $\mathbb{P}$-almost all $\omega\in\Omega$ the function

$$(\kappa,\theta,\gamma)\to X_t(\kappa,\theta,\gamma)$$

is continuously differentiable. For notational reasons we suppress the variables κ,θ and γ in most of the notation, with exception of the coefficient functions. Nevertheless one should always keep in mind, that all processes occurring are considered as functions in κ,θ and γ. Let us now assume that for all ϵ sufficiently small and all positive κ and θ as well as $\gamma\in\mathbb{R}^2$, the process $(Z_t(\epsilon,\kappa,\theta,\gamma))$ defined by

$$Z_t(\epsilon,\kappa,\theta,\gamma) = \mathcal{E}\left(-\int_0^t \sigma(X_s,\gamma))^{-1}\begin{pmatrix}0\\ \epsilon(\theta - v_s(\kappa,\theta,\gamma))\end{pmatrix}d\mathbb{W}_s\right),$$

is a martingale. Here $\mathcal{E}$ denotes the stochastic exponential (also known as the Doléans-Dade exponential see [19], II.8 page 85). Under the assumption that the coefficient function σ is uniformly coercitive, Z_t is always a martingale, if not, this must be checked in each case individually by using the Novikov condition. We assume at this point, that the function g is bounded away from

zero. The process $(Z_t(\epsilon,\kappa,\theta,\gamma))$ is precisely the Girsanov density process changing the law of $(X_t(\kappa,\theta,\gamma))$ to the law of $(X_t(\kappa+\epsilon,\theta,\gamma))$. In particular, we have for any Borel-measurable function $h:\mathbb{R}\to\mathbb{R}$ the following identity:

$$\mathbb{E}\left[h(X_T(\kappa+\epsilon,\theta,\gamma))\right]=\mathbb{E}\left[h(X_T(\kappa,\theta,\gamma))Z_T(\epsilon,\kappa,\theta,\gamma)\right].$$

Computing $Z_T(\epsilon,\kappa,\theta,\gamma)$ yields

$$Z_T(\epsilon,\kappa,\theta,\gamma) = \exp\left(\epsilon\left(\int_0^T\frac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1-\int_0^T\frac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2 -\frac{\epsilon}{2}\int_0^T\frac{(\theta-v_t)^2}{(1-\gamma_1^2)g(v_t,\gamma_2)^2}dt\right)\right)$$

Furthermore, we have for all $\epsilon\neq 0$ that

$$\begin{aligned}&\tfrac{1}{\epsilon}\left(Z_T(\epsilon,\kappa,\theta,\gamma)-1\right)=\tfrac{1}{\epsilon}\int_0^T dZ_t(\epsilon,\kappa,\theta,\gamma)\\ =\;&\tfrac{1}{\epsilon}\int_0^T Z_t(\epsilon,\kappa,\theta,\gamma)\cdot\epsilon\cdot\left(\tfrac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1-\tfrac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2\right)\\ =\;&\int_0^T Z_t(\epsilon,\kappa,\theta,\gamma)\left(\tfrac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1-\tfrac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2\right)\end{aligned}$$

Since g is bounded away from zero and $Z_0(\epsilon,\kappa,\theta,\gamma)=1$ we have

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(Z_T(\epsilon,\kappa,\theta,\gamma)-1\right)=\int_0^T\frac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1-\int_0^T\frac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2$$

as a limit in $L^2(\omega)$. Similarly, the process $(\tilde{Z}_t(\epsilon,\kappa,\theta,\gamma))$ defined by

$$\begin{aligned}&\tilde{Z}_t(\epsilon,\kappa,\theta,\gamma)=\mathcal{E}\left(-\int_0^t\sigma(X_s,\gamma)^{-1}\begin{pmatrix}0\\ \kappa\epsilon\end{pmatrix}d\mathbb{W}_s\right)\\ &=\exp\left(\epsilon\left(\int_0^T\frac{\gamma_1\kappa}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1-\int_0^T\frac{\kappa}{g(v_t,\gamma_2)}dW_t^2\right.\right.\\ &\left.\left.-\frac{\epsilon}{2}\int_0^T\frac{\kappa^2}{(1-\gamma_1^2)g(v_t,\gamma_2)^2}dt\right)\right)\end{aligned}$$

is the Girsanov density process changing the law of $(X_t(\kappa,\theta,\gamma))$ to the law of $(X_t(\kappa,\theta+\epsilon,\gamma))$ and in the same way as before, one obtains

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\tilde{Z}_T(\epsilon,\kappa,\theta,\gamma)-1\right) = \int_0^T \frac{\gamma_1\kappa}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1 - \int_0^T \frac{\kappa}{g(v_t,\gamma_2)}dW_t^2$$

The following proposition shows how to compute the derivatives of the option prices $\Pi(\kappa,\theta,\gamma)(h(S_T))$ with respect to κ and θ.

Proposition 4.1. *Under our general assumptions, we have for any Borel measurable function h s.t. $h(S_T(\kappa,\theta,\gamma)) \in L^2(\Omega)$ for all κ,θ and γ*

$$\frac{\partial}{\partial\kappa}\Pi(\kappa,\theta,\gamma)(h(S_T)) = \mathbb{E}\left[h(S_T)\left(\int_0^T \frac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1 - \int_0^T \frac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2\right)\right]$$

and

$$\frac{\partial}{\partial\theta}\Pi(\kappa,\theta,\gamma)(h(S_T)) = \mathbb{E}\left[h(S_T)\left(\int_0^T \frac{\gamma_1\kappa}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1 - \int_0^T \frac{\kappa}{g(v_t,\gamma_2)}dW_t^2\right)\right]$$

Proof. The result follows more or less directly from the discussion above. We have

$$\begin{aligned}
&\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\Pi(\kappa+\epsilon,\theta,\gamma)(h(S_T)) - \Pi(\kappa,\theta,\gamma)(h(S_T))\right]\\
=\ &\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\mathbb{E}(h(S_T(\kappa+\epsilon,\theta,\gamma))) - \mathbb{E}(h(S_T(\kappa,\theta,\gamma)))\right]\\
=\ &\lim_{\epsilon\to 0}\mathbb{E}\left(\frac{1}{\epsilon}\left(Z_T(\epsilon,\kappa,\theta,\gamma)-1\right)h(S_T(\kappa,\theta,\gamma))\right)\\
=\ &\mathbb{E}\left(\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(Z_T(\epsilon,\kappa,\theta,\gamma)-1\right)h(S_T(\kappa,\theta,\gamma))\right)\\
=\ &\mathbb{E}\left(\left(\int_0^T \frac{\gamma_1(\theta-v_t)}{\sqrt{1-\gamma_1^2}g(v_t,\gamma_2)}dW_t^1 - \int_0^T \frac{(\theta-v_t)}{g(v_t,\gamma_2)}dW_t^2\right)h(S_T(\kappa,\theta,\gamma))\right).
\end{aligned}$$

To interchange the order of taking the limit and expectation above, we used the L^2-convergence of the limit as well as the assumptions that $h(S_T(\kappa,\theta,\gamma)) \in L^2(\Omega)$. This proves that the partial derivative of $\Pi(\kappa,\theta,\gamma)(h(S_T))$ with respect to κ exists and is given by the expression stated in the proposition. The same argument works for the partial derivative of $\Pi(\kappa,\theta,\gamma)(h(S_T))$ with respect to θ by using the density $\tilde{Z}_T(\epsilon,\kappa,\theta,\gamma)$ defined in the discussion above. □

Since Girsanov transformations do not affect the volatility coefficients, the method just discussed does not help to compute the partial derivative of the option prices $\Pi(\kappa,\theta,\gamma)(h(S_T))$ with respect to γ. This derivative however, can be computed by using Proposition 3.3. For this let us define processes (A_t) and (B_t) via

$$\begin{aligned} A_t &:= \int_0^t \sqrt{1-\gamma_1^2} f'(v_s) B_s dW_s^1 + \int_0^t \gamma_1 f'(v_s) B_s dW_s^2 - \int_0^t f(v_s) f'(v_s) B_s ds \\ B_t &:= \exp(-\kappa t) \exp\left(\int_0^t \frac{\partial}{\partial x_2} g(v_s,\gamma_2) dW_s^2 - \int_0^t \frac{1}{2} \frac{\partial}{\partial x_2} g(v_s,\gamma_2)^2 ds \right), \end{aligned}$$

Again, these processes depend explicitly on the parameters κ, θ and γ, but we avoid the extensive notation $A_t(\kappa,\theta,\gamma)$ respectively $B_t(\kappa,\theta,\gamma)$ for the purpose of improving readability of the formulas. With this notation one can easily verify that the first variation process of X_t is given by

$$D^x X_t = \begin{pmatrix} \frac{S_t}{x} & S_t \cdot A_t \\ 0 & B_t \end{pmatrix}.$$

We define the $\mathbb{R}^{2\times 2}$-valued adapted process (M_t) via

$$M_t := \frac{1}{T} \begin{pmatrix} \frac{1}{\sqrt{1-\gamma_1^2} f(v_t) x} & \frac{A_t g(v_t,\gamma_2) - \gamma_1 B_t f(v_t)}{\sqrt{1-\gamma_1^2}} \\ 0 & \frac{B_t}{g(v_t,\gamma_2)} \end{pmatrix}$$

A straightforward computation shows $M_t = \frac{1}{T}\sigma(X_t,\gamma)^{-1} D^x X_t$. Furthermore, we define the following $\mathcal{F}_T$-measurable random variables

$$C_T := \int_0^T \gamma_1 x f'(v_s) - \frac{x A_s \frac{\partial}{\partial x_2} g(v_s,\gamma_2)}{B_s} \frac{\partial}{\partial \gamma_2} g(v_s,\gamma_2) ds$$

as well as K_T^j and L_T^j for $j = 1, 2$ via

$$K_T^1 := \begin{pmatrix} -2\gamma_1 x \int_0^T f(v_s)^2 ds & 0 \\ 0 & 0 \end{pmatrix}$$

$$K_T^2 := \begin{pmatrix} \gamma_1 x \int_0^T f(v_s)^2 ds & C_T \\ 0 & 0 \end{pmatrix}$$

$$L_T^1 := \begin{pmatrix} -2x \frac{\gamma_1}{\sqrt{1-\gamma_1^2}} \int_0^T f(v_s) dW_s^1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$L_T^2 := \begin{pmatrix} x\int_0^T f(v_s)dW_s^2 & -\int_0^T A_s x \frac{\partial}{\partial\gamma_2} g(v_s,\gamma_2)dW_s^2 \\ 0 & \int_0^T \frac{\frac{\partial}{\partial\gamma_2} g(v_s,\gamma_2)}{B_s} dW_s^2 \end{pmatrix}$$

In the situation of this section with the notation from above, the random variable Γ_T in Proposition 3.3 is precisely given by

$$\Gamma_T = L_T^1 + L_T^2 - (K_T^1 + K_T^2).$$

We obtain the following proposition.

Proposition 4.2. *With the notation from above, assume that (M_t) and Γ_T satisfy the assumptions in Proposition 3.3 and Corollary 3.1. Let $h \in L^2_{loc}(\mathbb{R},\mathbb{R})$ with at most linear growth at infinity s.t. $h(S_T(\kappa,\theta,\gamma)) \in L^2(\Omega)$ for all κ,θ and γ, then*

$$\left(\frac{\partial}{\partial\gamma_1}\Pi(\kappa,\theta,\gamma)(h(S_T)), \frac{\partial}{\partial\gamma_2}\Pi(\kappa,\theta,\gamma)(h(S_T)\right) = \mathbb{E}(h(S_T(\kappa,\theta,\gamma))\cdot\rho_T)$$

for the (1×2) matrix valued weight function ρ_T given by

$$\rho_T := \delta(M_t\cdot\Gamma_T)$$

Proof. The result follows from Proposition 3.3 after solving all the linear stochastic differential equations for the first variation process and performing some (2×2)-matrix calculus. That the integrand in the Skorohod integral is indeed in the domain of δ follows from our assumptions on the market models. □

The expression $\delta(M_t\cdot\Gamma_T)$ in the proposition above, can be computed using Lemma 3.1

5 The Malliavin gradient method for calibration

As discussed in section 2, calibrating our mean reverting stochastic volatility model means picking out that model of our family, which has the property that the corresponding parameters κ,θ and γ minimize a certain functional. This functional should measure the distance of theoretical prices based on the model to observed market prices. There are several possibilities to choose this functional. The most natural is the one discussed in section 2

$$L(\kappa,\theta,\gamma) = \sum_{i=1}^{n} \left[\Pi(\kappa,\theta,\gamma,)(h_i(S_{T_i})) - \Pi^{obs}(h_i(S_{T_i}))\right]^2$$

where $h_i(S_{T_i})$ are European style options and $\Pi^{obs}(h_i(S_{T_i}))$ are observed market prices corresponding to these options. Let us now use gradient methods to find minimizers for the functional above. The partial derivatives of L with respect to κ, θ and γ are as follows : With $D_i(\kappa,\theta,\gamma) = \Pi(\kappa,\theta,\gamma,)(h_i(S_{T_i})) - \Pi^{obs}(h_i(S_{T_i}))$ one has

$$\begin{aligned}
\frac{\partial}{\partial\kappa}L(\kappa,\theta,\gamma) &= 2\sum_{i=1}^{n} D_i(\kappa,\theta,\gamma)\cdot\frac{\partial}{\partial\kappa}\Pi(\kappa,\theta,\gamma)(h_i(S_{T_i})) \\
\frac{\partial}{\partial\theta}L(\kappa,\theta,\gamma) &= 2\sum_{i=1}^{n} D_i(\kappa,\theta,\gamma)\cdot\frac{\partial}{\partial\theta}\Pi(\kappa,\theta,\gamma)(h_i(S_{T_i})) \\
\frac{\partial}{\partial\gamma_j}L(\kappa,\theta,\gamma) &= 2\sum_{i=1}^{n} D_i(\kappa,\theta,\gamma)\cdot\frac{\partial}{\partial\gamma_j}\Pi(\kappa,\theta,\gamma)(h_i(S_{T_i}))
\end{aligned}$$

for $j = 1, 2$. The computation of $D_i(\kappa,\theta,\gamma)$ only involves the computation of the option prices in the corresponding model. If closed form solutions exist we take them. If not we use Monte Carlo Simulation. For the computation of the partial derivatives $\frac{\partial}{\partial\kappa}\Pi(\kappa,\theta,\gamma)(h_i(S_{T_i}))$, $\frac{\partial}{\partial\theta}\Pi(\kappa,\theta,\gamma,)(h_i(S_{T_i}))$ and $\frac{\partial}{\partial\gamma_j}\Pi(\kappa,\theta,\gamma,)(h_i(S_{T_i}))$ we use the formulas computed in section 4. To compute these formulas we also use Monte Carlo simulation. To simulate the process (X_t) describing the stock price and the volatility we use general methods for simulating solutions of stochastic differential equations like the stochastic Euler-scheme or the Milstein-scheme (see [16]). Always when Monte Carlo methods are used, one can think about variance reduction. In general, one can say that all the general theory like importance sampling, stratified sampling and the methods discussed in [7] which particularly suit the Malliavin calculus can be applied. When using gradient methods with step-size control, it can be said that for computation of the step-size one also has to use Monte Carlo Simulation, but in general one can run these with a much lower number of simulations of the process (X_t). We will discuss the case of the Vasicek model in Appendix A. We have chosen this model, because it is one of the simplest stochastic volatility models and the formulas obtained in the last section turn out to be of reasonable length. Corresponding formulas for other stochastic volatility models turn out to be quite lengthy instead. The case of the Heston model, which is certainly one of the most popular stochastic volatility models deserves special attention. One of

the difficulties when dealing with the Heston model comes from the fact that the volatility of the volatility is a square-root function and therefore does not satisfy the technical conditions in the previous section. One can avoid these difficulties by using a multidimensional model and a drift adjustment technique such as in [3]). On the other side, it has recently be proved that the Heston volatility is indeed Malliavin differentiable, see [5]. The localization techniques in this article combined with the techniques from the last section in fact prove, that Proposition 4.1 and proposition 4.2 remain valid in the context of the Heston model. The same holds for the case of the Hull and White model. The formulas given in section 4 can easily be extended to cover second order derivatives and the Hessian of $\Pi(\kappa, \theta, \gamma,)(h_i(S_{T_i}))$ so that one can implement the Newton method as discussed in Appendix B.

6 Conclusions

The method proposed in the article presents the new idea, to combine Malliavin calculus and standard methods from Optimization, in order to calibrate stochastic volatility models. The method is presented in a way, that it can easily be adapted to the special needs of financial engineers working in practice on individual problems. It has been tested successfully in cooperation with several institutes at the ITWM Fraunhofer Institute in Kaiserslautern. There is open space to optimize the method, by using standard tricks to fasten up Monte Carlo methods. The Malliavin Gradient Method appears to be a real alternative to classical calibration methods from numerical partial differential equations and statistics. We should also mention that the formulas for the derivatives obtained in this article can be used to get good estimates for Value at Risk in stochastic volatility models. We will present this interesting application in a future publication.

Appendix A : Examples and simulations

We consider the case of a family of Vasicek models given by the family of stochastic differential equations

$$\begin{aligned} dS_t &= S_t\left(\sqrt{1-\gamma_1^2}v_t dW_t^1 + \gamma_1 v_t dW_t^2\right) \\ dv_t &= \kappa(\theta - v_t)dt + \gamma_2 dW_t^2 \end{aligned}$$

for positive parameters κ and θ and $\gamma \in \mathbb{R}^2$. As in the previous section we write $\Pi(\kappa, \theta, \gamma)(h(S_T))$ for the price of a European style option $h(S_T)$ in

this model, where h is assumed to satisfy the technical assumptions made in section 4. In the notation of section 1 we have to chose $f(x_2) = x_2$ and $g(x_2, \gamma_2) = \gamma_2$. This choice results in the following derivatives

$$f'(x_2) = 1, \quad \frac{\partial}{\partial x_2} g(x_2, \gamma_2) = 0, \quad \frac{\partial}{\partial \gamma_2} = 1.$$

It follows from Proposition 4.1. that

$$\begin{aligned} \frac{\partial}{\partial \kappa} \Pi(\kappa, \theta, \gamma)(h(S_T)) &= \mathbb{E}\left[h(S_T) \cdot \left(\frac{\gamma_1}{\gamma_2 \sqrt{1-\gamma_1^2}} \left(\theta W_T^1 - \int_0^T v_t dW_t^1 \right) \right.\right. \\ &\quad \left.\left. - \frac{1}{\gamma_2} \left(\theta W_T^2 - \int_0^T v_t dW_t^2 \right) \right) \right] \end{aligned}$$

as well as

$$\frac{\partial}{\partial \theta} \Pi(\kappa, \theta, \gamma)(h(S_T)) = \mathbb{E}\left[h(S_T) \cdot \frac{\kappa}{\gamma_2} \left(\frac{\gamma_1}{\sqrt{1-\gamma_1}^2} W_T^1 - W_T^2 \right) \right].$$

The processes (A_t) and (B_t) defined in preparation for Proposition 4.2. can easily be computed as

$$\begin{aligned} A_t &= \sqrt{1-\gamma_1^2} \int_0^t e^{-\kappa s} dW_s^1 + \gamma_1 \int_0^t e^{-\kappa s} dW_s^2 - \int_0^t v_s e^{-\kappa s} ds \\ B_t &= e^{-\kappa t} \end{aligned}$$

It then follows that

$$\frac{\partial}{\partial \gamma_2} \Pi(\kappa, \theta, \gamma)(h(S_T)) = \mathbb{E}\left[h(S_T) \frac{1}{\gamma_2} \left(\left(\int_0^T e^{\kappa s} dW_s^2 \right) \left(\int_0^T e^{-\kappa s} dW_s^2 \right) + T \right) \right].$$

The formula for $\frac{\partial}{\partial \gamma_1} \Pi(\kappa, \theta, \gamma)(h(S_T))$ is quite lengthy and rather suited for typing into a computer program than in a scientific article. We therefore omit it here and assume in the following that the correlation parameter γ_1 is given and set up a gradient method involving the remaining parameters κ, θ and γ_2. This assumption has been made, also to provide a nice three dimensional rather than 4 dimensional graphical output. We use the techniques presented in section 5 using the formulas above for the partial derivatives of $\Pi(\kappa, \theta, \gamma)(h(S_T))$. The setup is now as follows. We consider three digital

call options with strike prices 0.8, 1.0 and 1.2 respectively and maturity time $T = 1$, i.e.

$$\begin{aligned} h_1(S_T) &= \mathbf{1}_{\{S_1 \geq 0.8\}} \\ h_2(S_T) &= \mathbf{1}_{\{S_1 \geq 1.0\}} \\ h_3(S_T) &= \mathbf{1}_{\{S_1 \geq 1.2\}}. \end{aligned}$$

We assume that the real market parameters are given as

$$\begin{aligned} \kappa &= 0.5 \\ \theta &= 0.2 \\ \gamma_1 &= -0.2 \\ \gamma_2 &= 0.1 \end{aligned}$$

and that the initial price of the stock is $S_0 = 1$ whereas the initial volatility is $v_0 = 0.2$. Prices for the digital calls $h_1(S_T), h_2(S_T)$ and $h_3(S_T)$ can be computed by Monte Carlo simulation. The results of our simulation, where we used $50 * 10^6$ simulated paths of (S_t) and (v_t) each with a time discretization of $\Delta = 0.01$, are shown in the following table :

Table 1:

strike	price
0.8	0.81909648
1.0	0.43567714
1.2	0.13018866

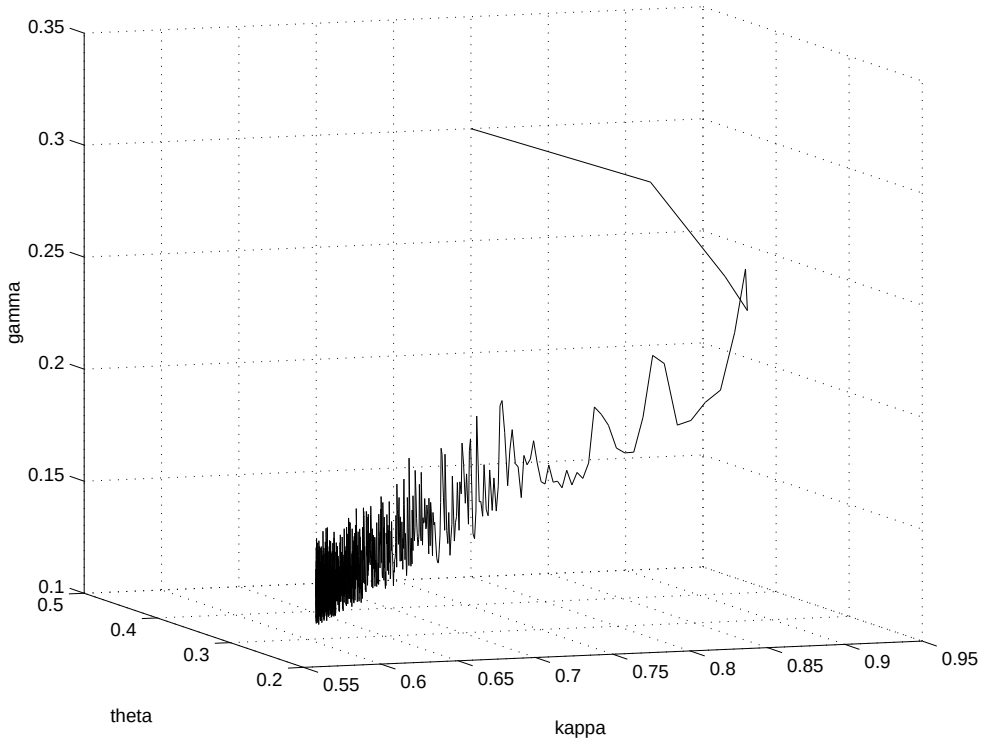
The trader at the market observes these prices, but does not know about the values of κ, θ and γ_2. He does know though that the market is given by a Vasicek model and that the covariation γ_1 between the two driving Brownian motions is $\gamma_1 = -0.2$. He can then start Monte Carlo simulations by himself, trying various parameters he might be lucky and get close to the observed market prices. More systematically he can try the gradient method proposed in the previous section in order to get some idea where to look for a good set of parameters κ, θ, γ_2. The results of such a gradient method, where we used 10^5 simulated paths for the computation of the gradient with a time-discretization of $\Delta = 0.01$ in each step, as well as 1000 simulations for the step size control and a step size discretization of 0.0025 with maximum step size 0.05, are shown in the following table and the figure on the next page.

Table 2:

	initial	250	500	750	1000
κ	0.8	0.6164	0.5912	0.5792	0.5718
θ	0.5	0.2468	0.2371	0.2320	0.2290
γ_2	0.3	0.1431	0.1480	0.1283	0.1273

The first row in the table shows the number of steps performed by the gradient method. Simulations were done under $Matlab^{©}$ using a PC with an $AMDAthlon64^{©}$ processor. The computing time for this simulation was approximately four hours. The results in the table show that in the end the convergence becomes very slow and a considerable higher computational effort has to be made to get closer to the real parameters. A closer look at the figure also shows that the convergence of the γ_2 component is affected by the high variance of the corresponding weight in the formula for $\frac{\partial}{\partial\gamma_2}\Pi(\kappa,\theta,\gamma)(h(S_T))$. The algorithm can be significantly improved by using general variance reduction methods as mentioned in the introduction.

Figure 1:



Appendix B : Gradient methods

In this appendix we present the general idea behind gradient methods for solving minimization problems. We do not go into a detailed discussion of the numerics (for this see the abundant literature on optimization, for example [6], [2]) but rather illustrate the algorithm. Consider the unconstrained problem:

- given a function $f : \mathbb{R}^n \to \mathbb{R}$ find a minimizer $x^* \in \mathbb{R}^n$ for f.

We assume that the function f is at least of class C^2. Under these assumptions a necessary condition for x^* minimizing f is that the gradient of $\nabla f(x^*)$ vanishes. Given an arbitrary point $x_0 \in \mathbb{R}^n$ and a direction $\gamma \in \mathbb{R}^n$, the function f at x_0 increases in direction γ, if the scalar product $(\nabla f(x_0), \gamma)$ is greater than 0 and decreases in direction γ, if $(\nabla f(x_0), \gamma)$ is less than 0. In the latter case γ is called a direction of descent. If x_0 is already a minimizer we stop. Otherwise, assuming that x_0 is no saddle point, we choose a direction γ_0 of descent, i.e. $\gamma_0 \in \mathbb{R}^n$ such that

$$(\nabla f(x_0), \gamma_0) < 0.$$

We then choose a step-size t_0 and make a step from x_0 into direction γ_0 of size t_0. Then we arrive at $x_1 := x_0 + t_0 \cdot \gamma_0$. If x_1 is a (local) minimizer we stop. If not we go on, chose a direction γ_1 of descent and a step-size t_1 in order to get x_2 and so on. Inductively :

- find γ_k s.t. $(\nabla f(x_k), \gamma_k) < 0$, choose a step-size t_k and set

$$x_{k+1} = x_k + t_k \cdot \gamma_k.$$

In an iterative procedure we may hope to find a minimizer for f or at least to get close to one. Obviously the question remains how to choose the direction of descent γ_k and the step-size t_k in each iterative step. One possibility, is to choose

$$\gamma_k = -\nabla f(x_k).$$

Then obviously

$$(\nabla f(x_k), \gamma_k) = (\nabla f(x_k), -\nabla f(x_k)) = -\|\nabla f(x_k)\| \leq 0.$$

A natural idea is to incorporate not only first order data but also second order data, i.e. the curvature of the function f. In fact, for any symmetric positive definite matrix H_k one has

$$(\nabla f(x_k), -H_k \cdot \nabla f(x_k)) < 0$$

and therefore one can choose for the descent directions the vectors

$$\gamma_k = -H_k \cdot \nabla f(x_k).$$

If one chooses H_k to be the inverse of the Hessian of f at x_k and chooses the step-size to be constant one speaks of the Newton method (see Satz 3.1. section 7.3, page 192 in [22] for convergence of this method). Often however, the computational effort to compute the Hessian at every iterative step is to big and one takes H_k to be some approximation of the Hessian. Better convergence results can be obtained by an intelligent step-size control. In principle, the following idea helps in determining the right step-size :

- at step k take $t_k > 0$ to be the first local minimizer of the function

$$t \mapsto f(x_k + t \cdot \gamma_k)$$

If the function f is not given explicitly by a formula and the evaluation of f at some point leads to computational effort which is not negligible, this idea leads to difficulties. What one can do and what we do in the setup of the Malliavin gradient method is the following :

- choose a maximum step-size $t_{\max}$ and a grid constant $m \in \mathbb{N}$ at the beginning and at each iterative step choose l^* s.t.

$$f\left(x_k + \frac{l^*}{m} t_{\max} \cdot \gamma_k\right) = \min\left\{ f\left(x_k + \frac{l}{m} t_{\max} \cdot \gamma_k\right) | l = 1, .., m \right\}$$

  and set $t_k = \frac{l^*}{m} t_{\max}$.

For a convergence result with variable step-size see [22], Satz 2.5, Kapitel 7.2.2. More advanced methods do not only use information about the gradient at the k-th step but also use information on the previous directions. For example, one can determine the direction in the k-th step via the formula

$$\gamma_k = -\nabla f(x_k) + \beta_k \cdot \gamma_{k-1}.$$

where β_k is a scalar which determines the portion of the previous direction to be added to determine the new direction. The following choices for β_k are often used :

- Fletcher-Reeves formula :

$$\beta_k = \frac{< \nabla f(x_k), \nabla f(x_k) >}{< \nabla f(x_{k-1}), \nabla f(x_{k-1}) >}$$

- Polak-Riebiere formula :

$$\beta_k = \frac{< \nabla f(x_{k-1}) - \nabla f(x_k), \nabla f(x_k) >}{< \nabla f(x_{k-1}), \nabla f(x_{k-1}) >}$$

Appendix C : A brief review on Malliavin calculus

In this appendix, we summarize the construction of the Malliavin derivative operator and shortly revise its main properties. Though changed by its appearance, the material presented here has mainly been taken from the classical references [18] and [21].

Assume we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which there is defined an m-dimensional Brownian motion $\mathbb{W}$. We would like to differentiate functionals of the form

$$F : \Omega \to \mathbb{R}$$

or at least those of a certain nice subclass. For this we first assume that the functional is given by

$$F = f\left(\mathbb{W}_{t_1}, ..., \mathbb{W}_{t_l}\right)$$

where $f \in C_b^\infty\left((\mathbb{R}^m)^l\right)$ is a smooth function with bounded derivatives of all orders. Given $h \in L^2([0,T], \mathbb{R}^m)$ we have that $\int_0^\cdot h(s)ds \in C_0([0,T], \mathbb{R}^m)$ where the integral is computed component wise and the dot indicates that the upper bound of the integral is taken as a variable. The subspace of $C_0([0,T], \mathbb{R}^m)$ generated by this kind of functions is called the Cameron-Martin space. The directional derivative of F in direction $\int_0^\cdot h(s)ds$ at ω is given by

$$\begin{aligned} D_h F(\omega) &:= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \tilde{F}\left(\mathbb{W}(\omega) + \epsilon \cdot \int_0^\cdot h(s)ds\right) \\ &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} f\left(\mathbb{W}_{t_1}(\omega) + \int_0^{t_1} h(s)ds, ..., \mathbb{W}_{t_l}(\omega) + \int_0^{t_l} h(s)ds\right) \\ &= \sum_{i=1}^m \nabla_i f\left(\mathbb{W}_{t_1}(\omega), ..., \mathbb{W}_{t_l}(\omega)\right)^\top \cdot \int_0^{t_i} h(s)ds. \end{aligned}$$

where $\tilde{F}$ is the functional on the m-dimensional Wiener space through which F factorizes and ∇_i denotes the gradient with respect to the i-th m-dimensional argument in f. Now for fixed ω consider the linear bounded functional on $L^2([0,T],\mathbb{R}^m)$ given by

$$h \mapsto D_h F(\omega).$$

By the Riesz-representation theorem there is an element $DF(\omega)$ in $L^2([0,T],\mathbb{R}^m)$ which is considered as a row vector such that

$$D_h F(\omega) \;=\; < h, DF(\omega)^\top >_{L^2([0,T],\mathbb{R}^m)} = \int_0^T DF(\omega)(s)h(s)ds$$

$\forall h \in L^2([0,T],\mathbb{R}^m)$. In the following we denote $DF(\omega)(s)$ with $D_s F(\omega)$. Let us now consider ω as a variable. The assumption that f has bounded derivatives of all orders ensures that for all $p \geq 1$ we have $DF \in L^p\left(\Omega, L^2([0,T],\mathbb{R}^m)\right)$ when considered as an $L^2([0,T],\mathbb{R}^m)$ valued functional in ω

Assume now that the functional F is not necessarily cylindrical but there exists a sequence of cylindrical functionals F_i such that (F_i) converges to F in $L^p(\Omega)$ and (DF_i) converges to G in $L^p(\Omega, L^2([0,T],\mathbb{R}^m))$. Then we define

$$DF := G = \lim_{i\to\infty} DF_i.$$

Using the Cameron-Martin Theorem it is not hard to show that if one has another sequence $(\tilde{F}_i)$ converging to F in $L^p(\Omega)$ such that $(D\tilde{F}_i)$ converges to $\tilde{G}$ in $L^p(\Omega, L^2([0,T],\mathbb{R}^m))$, then $G = \tilde{G}$ in $L^p(\Omega, L^2([0,T],\mathbb{R}^m))$. This basically shows that the operator

$$D : L^p(\Omega) \to L^p(\Omega, L^2([0,T],\mathbb{R}^m))$$

defined on the cylindrical functionals is closable.

Definition C 1. *For $p \geq 1$ we define the Malliavin derivative operator*

$$D : L^p(\Omega) \to L^p(\Omega, L^2([0,T],\mathbb{R}^m))$$

as the closure of the operator above. For F in the domain of D we define

$$\|F\|_{1,p} := \|F\|_{L^p(\Omega)} + \|DF\|_{L^p(\Omega,(L^2[0,T],\mathbb{R}^m))}.$$

Then the domain of D is precisely the closure of the cylindrical functionals under the norm above. It will be denoted with $\mathbb{D}_{1,p}$.

If the functional F is vector-valued then the Malliavin derivative is computed component wise and considered as a Matrix in the same way as the Jacobian matrix in standard calculus. In the following, we restrict ourselves to the case where $p = 2$. Then we are dealing with Hilbert spaces.

Definition C 2. *The adjoint operator $\delta = D^*$*

$$\delta : L^2(\Omega \times [0,T], \mathbb{R}^m) \to L^2(\Omega)$$

is called the Skorohod integral. We denote its domain with $dom(\delta)$.

The word "integral" is motivated by the following proposition.

Proposition C 1. *The class $L^2_a(\Omega \times [0,T], \mathbb{R}^m)$ of adapted square integrable processes is contained in $dom(\delta)$ and on this class the Skorohod integral coincides with the Itô-integral.*

The following formula is called the integration by parts formula of Malliavin calculus.

Proposition C 2. *If $F \in \mathbb{D}_{1,2}$ and $u \in dom(\delta)$ then*

$$\mathbb{E}\left(\int_0^T D_t F \cdot u(\cdot,t)dt\right) = \mathbb{E}(F \cdot \delta(u)). \tag{1}$$

Proof. This follows directly from the definition of δ as the adjoint operator of D. □

Another useful formula for computing Malliavin derivatives is the following chain-rule (see [17], Lemma 2.1) :

Proposition C 3. *Let $\phi : \mathbb{R}^k \to \mathbb{R}$ be a continuously differentiable function and $F = (F_1, ..., F_k)$ such that $F_i \in \mathbb{D}_{1,2}$. Then $\phi(F) \in \mathbb{D}_{1,2}$ if and only if $\phi(F) \in L^2(\Omega)$ and $\nabla\phi(F)^\top DF \in L^2(\Omega \times [0,T], \mathbb{R}^m)$ and in this case*

$$D_t\phi(F) = \nabla\phi(F)^\top \cdot D_t F.$$

References

[1] Bellini, F.; Frittelli, M. On the existence of minimax martingale measures. Math. Finance 12, No.1, 1-21 (2002).

[2] Bonnans, J.F.; Gilbert, J.C.; Lemaréchal, C.; Sagastizábal, C.A. Numerical optimization. Theoretical and practical aspects. Transl. from the French. Universitext. Berlin: Springer (1997).

[3] Ewald, C.-O. Local Volatility in the Heston Model : A Malliavin Calculus Approach. Journal of Applied Mathematics and Stochastic Analysis, Vol. 2005 (3)

[4] Ewald, C.-O. Optimal logarithmic Utility and optimal Portfolios for an Insider in a Stochastic Volatility Model. International Journal of Theoretical and Applied Finance Vol. 8 No. 3 (2005)

[5] Ewald, C.-O.; Alos, E. A Note on the Malliavin differentiability of the Heston Volatility. To appear in Finance Stoch.

[6] Fletcher, R. Practical methods of optimization. Repr. of the 2nd ed. Chichester: Wiley (2001)

[7] Fournié, E.; Lasry, J.-M.; Lebuchous, J.; Lions, P.-L. Applications of Malliavin calculus to Monte Carlo methods in finance. Finance Stoch. 3, No.4, 391-412 (1999).

[8] Fouque, J.-P.; Papanicolaou, G.; Sircar, K.R. Derivatives in financial markets with stochastic volatility. Cambridge: Cambridge University Press (2000).

[9] Fujiwara, T.; Miyahara, Y. The minimal entropy martingale measures for geometric Lvy processes. Finance Stoch. 7, No.4, 509-531 (2003)

[10] Glasserman, Paul Monte Carlo methods in financial engineering. Applications of Mathematics 53. New York, NY: Springer (2004).

[11] Goll, Thomas; Rüschendorf, Ludger Minimax and minimal distance martingale measures and their relationship to portfolio optimization. Finance Stoch. 5, No.4, 557-581 (2001).

[12] Goll, Thomas; Rüschendorf, Ludger. Minimal distance martingale measures and optimal portfolios consistent with observed market prices. (preprint)

[13] Heston, S.L.; A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options; Review of Financial Studies, Vol.6, Issue 2 (1993)

[14] Hull, J.; White, A. The pricing of options on assets with stochastic volatilities. J. Finance 42(2) (1987)

[15] Ikeda, N.; Watanabe, S. Stochastic differential equations and diffusion processes. 2nd ed. North-Holland Mathematical Library, 24. Amsterdam etc.: North-Holland; Tokyo: Kodansha Ltd. xvi, 555 p. (1989).

[16] Kloeden, P.E.; Platen, E. Numerical solution of stochastic differential equations. Applications of Mathematics. 23. Berlin: Springer-Verlag (1992).

[17] León, J.A.; Navarro, R.; Nualart, David An anticipating calculus approach to the utility maximization of an insider. Math. Finance 13, No.1, 171-185 (2003).

[18] Nualart, D. The Malliavin calculus and related topics. Probability and Its Applications. New York, NY: Springer-Verlag (1995).

[19] Protter, P.E. Stochastic integration and differential equations. 2nd ed. Applications of Mathematics 21. Berlin: Springer (2004)

[20] Sanz-Sol'e M. A course on Malliavin calculus with applications to stochastic partial differential equations. Institut de Matemàtica Universitat de Barcelona, Lecture Notes vol.2 (2004)

[21] Üstünel, A.S. An introduction to analysis on Wiener space. Lecture Notes in Mathematics. 1610. Berlin: Springer-Verlag. x, 95 p.;(1995).

[22] Werner, Jochen Numerische Mathematik. Band 2: Eigenwertaufgaben, lineare Optimierungsaufgaben, unrestringierte Optimierungsaufgaben. Friedr. Vieweg & Sohn (1992).

[23] Yosida, K. Functional analysis. Berlin: Springer-Verlag (1994).