CALIBRATING VOLATILITY SURFACES VIA RELATIVE-ENTROPY MINIMIZATION

by

Marco Avellaneda, Craig Friedman, Richard Holmes and Dominick Samperi[1]

We present a framework for calibrating a pricing model to a prescribed set of option prices quoted in the market. Our algorithm yields an arbitrage-free diffusion process that minimizes the relative entropy distance to a prior diffusion. We solve a constrained (minimax) optimal control problem using a finite-difference scheme for a Bellman parabolic equation combined with a gradient-based optimization routine. The number of unknowns in the optimization step is equal to the number of option prices that need to be matched, and is independent of the mesh-size used for the scheme. This results in an efficient, non-parametric calibration method that can match an arbitrary number of option prices to any desired degree of accuracy. The algorithm can be used to interpolate, both in strike and expiration date, between implied volatilities of traded options and to price exotics. The stability and qualitative properties of the computed volatility surface are discussed, including the effect of the Bayesian prior on the shape of the surface and on the implied volatility smile/skew. The method is illustrated by calibrating to market prices of Dollar-Deutschemark over-the-counter options and computing interpolated implied-volatility curves.

Keywords: Option pricing, implied volatility surface, calibration, relative entropy, stochastic control, volatility smile and skew.

[1] Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY, 10012. This research was supported by the National Science Foundation.

1. Introduction

1.1 Deriving a diffusion model from option prices

It is well known that the constant-volatility assumption made in the Black-Scholes framework for option pricing is not valid in real markets. For example, S&P 500 index options are such that out-of-the money puts have higher implied volatilities than out-of-the money calls. In the currency options markets, implied volatilities exhibit a "smile" and a "skew" (in both maturity and strike) whereby at-the-money options trade at lower volatilities than other strikes, and a premium for puts in one of the two currencies is manifest in the price of "risk-reversals"[2]. To model the strike- and maturity-dependence of implied volatility, researchers have proposed using arbitrage-free diffusion models for the underlying index in which the spot volatility coefficient is a function of the index level and time. The problem is then to determine what this volatility "surface" should be, given the observed option prices.

This paper present a simple, rigorous, method for constructing such an arbitrage-free diffusion process. The basic idea is to assume an initial Bayesian prior distribution for the evolution of the index and to modify it to produce a calibrated model such that the corresponding probability is as close as possible to the prior. For this, we use the concept of Kullback-Leibler information distance, or relative entropy.

The basic approach is as follows. Let

$$\frac{dS_t}{S_t} = \sigma_t dZ_t \ + \ \mu \, dt \tag{1.1}$$

represent the process that we wish to determine. Here σ_t is a random process adapted to the standard information flow and μ is the risk-neutral drift, which we assume is known[3] . The calibration conditions for M traded options can be written as

$$\mathbf{E}^{\sigma}\left[e^{-rT_i} G_i(S_{T_i})\right] = C_i, \ i = 1, 2, \dots, M \ , \tag{1.2}$$

where r is the interest rate, $\mathbf{E}^{\sigma}[\cdot]$ denotes the expectation with respect to the measure corresponding to (1.1) and $G_i(S_{T_i}),\ C_i,\ i = 1, 2, \dots, M$ represent, respectively, the payoffs and prices of the M options that we wish to match.

We will show that minimizing relative entropy is essentially equivalent to minimizing the functional

[2] A risk-reversal is a position consisting in being long a call and short a put with symmetric strikes.

[3] μ is the interest rate differential (carry) in foreign exchange and the interest rate minus the dividend yield for equity indices. We assume therefore a "risk-adjusted" drift.

$$\mathbf{E}^{\sigma}\left[\int_0^T \eta(\sigma_s^2)ds\right] , \tag{1.3}$$

where $\eta(\sigma_s^2)$ is a strictly convex function which vanishes at the volatility of the prior distribution.

This constrained stochastic control problem is equivalent to a Lagrange multiplier problem in which we maximize the augmented objective function

$$\mathbf{E}^{\sigma}\left[-\int_0^T \eta(\sigma_s^2)ds + \sum_{i=1}^{M} \lambda_i\, e^{-rT_i} G_i(S_{T_i})\right] - \sum_{i=1}^{M} \lambda_i\, C_i, \tag{1.4}$$

over all adapted volatility processes σ_t and then minimize the result over $(\lambda_1, \ldots \lambda_M)$.[4]

We show that in the absence of arbitrage opportunities the value function $V(\lambda_1, \ldots \lambda_M)$ corresponding to (1.4) is smooth and strictly convex in λ. In particular, it has a unique minimum. The first-order condition at the minimum,

$$\frac{\partial V}{\partial \lambda_i} = \mathbf{E}^{\sigma^*}\left[e^{-rT_i} G_i(S_{T_i})\right] - C_i = 0,\ i = 1, 2, \ldots , M,$$

ensures that the model is calibrated to market prices. Hence, in this approach, *calibrating the model to the M option prices is equivalent to finding the minimum of a convex function of M variables.*

The algorithm for computing $V(\lambda_1, \ldots \lambda_M)$ for a given set of Lagrange multipliers consists in solving the Bellman partial differential equation corresponding to (1.4) *viz.*,

$$V_t + \frac{1}{2} e^{r\,t} \Phi\left(\frac{e^{-r\,t}}{2} S^2\, V_{SS}\right) + \mu\, S\, V_S - r\, V =$$

$$- \sum_{t\,<\,T_i}^{M} \lambda_i \left(G_i(S_{T_i}) - e^{r\,T_i}\, C_i\right) \delta(t - T_i) ,$$

where Φ is the Legendre dual of η. This is done numerically for successive choices of $(\lambda_1, \ldots \lambda_M)$ until the minimum of

[4] In practice, we shall restrict our search to volatility processes that satisfy uniform bounds $0 < \sigma_{\min} \leq \sigma_s \leq \sigma_{\max}$. This constraint will typically not be binding except in a neighborhood of points in the (S, t)-plane corresponding to each strike/expiration date.

$$V(\lambda_1, \dots \lambda_M) \;=\; V(S, 0; \lambda_1, \dots \lambda_M)$$

is reached. The optimal volatility surface is identified as

$$\sigma_t^* \;=\; \sigma(S,t) \;=\; \sqrt{\Phi'\left(\frac{e^{-rt}\, S^2\, V_{SS}(S,\,t)}{2}\right)}\;. \tag{1.5}$$

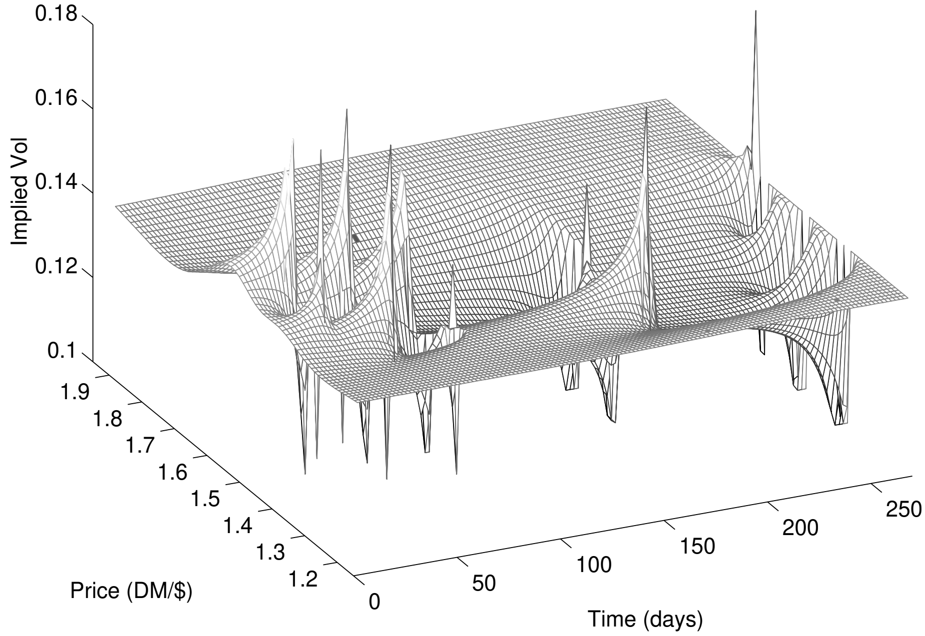

FIGURE 1. Calibrated volatility surface for a set of 25 options on Dollar-Deutchemark (dataset in Appendix B). The prior in this calculation is $\sigma_0 \;=\; 0.141$. The surface consists of "humps" and "troughs" originating near each strike/expiration date which are smooth away from these points. At the strike/expiration points, the volatility peaks are $\sigma_{min} \;=\; 0.10$ or $\sigma_{max} \;=\; 0.20$. Notice that the surface converges to the prior volatility away from the input strikes.

The volatility function thus obtained is what is traditionally called an "implied (spot) volatility surface".

A few remarks are in order. First, this approach permits the user to impose his or her preference ordering via the specification of a Bayesian prior: the diffusion selected by the model matches market prices and is also as close as possible to the prior. The specification of

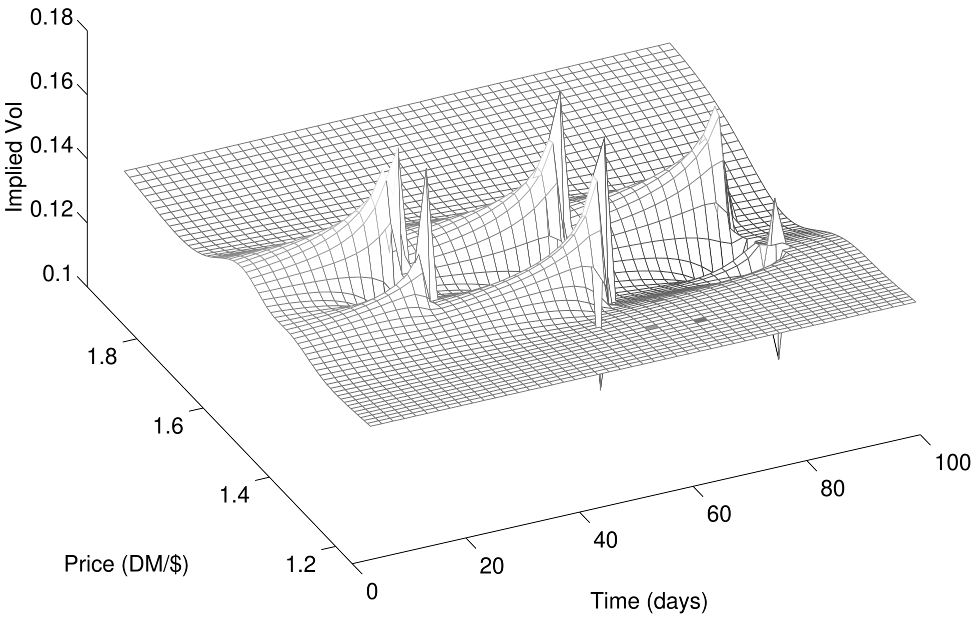

FIGURE 2. Detail of the volatility surface of Figure 1 corresponding to the first 100 days after the trading date. Notice that the price information corresponding to maturities after 90 days affects the earlier values of the surface at earlier dates, as a trough inherited from the later maturities appears in the period from 90 to 100 days.

a prior distribution is a key feature of the procedure.[5] Minimizing the relative entropy with respect to the prior stabilizes the far-tails of the probability distribution for the underlying index and implies smoothness of the volatility surface (1.5).[6] The procedure leads to a simple and numerically stable method for calibrating a pricing model. The small number of input parameters that need to be adjusted makes it tractable in practice. This is in contrast to other proposals where ad-hoc adjustments are required to achieve a stable algorithm.

Figures 1 and 2 display a calibrated spot volatility surface $\sigma(S, t)$ corresponding to a dataset corresponding to Dollar-Deutchemark over-the-counter options for the date of August 23, 1995, provided to us by a market-maker. It consisted of 25 option prices, corresponding to 20- and 25-delta puts and calls and 50-delta calls for maturities of 30, 60, 90, 180 and 270 days[7] . The model was calibrated to mid-market quotes to an accuracy of

[5] The prior volatility need not be a constant. It can be, for instance, a function of time and/or price.

[6] As we shall see, a unique prescription of the volatility surface far way from traded strikes cannot be obtained precisely from option prices. The introduction of the Bayesian prior serves as an "extrapolation mechanism" for characterizing the volatility in regions where the price information is weak, e.g. for strikes which are deeply away-from-the-money, as well as a mechanism for smoothing the volatility surface.

[7] A 25-delta put is a put with a Black-Scholes delta of -0.25, etc. This is standard terminology for over-the-counter currency options.

10^{-4} (in relative terms). The complete dataset is included in Appendix B.

Generically, the volatility surface corresponding to calibrating to a finite number of option prices converges to the prior volatility surface for (S, t) far away from strikes/expiration dates. Significant variations of the volatility surface occur near strikes/expiration dates. These distortions are sharp near the strikes/expiration dates and diffuse smoothly away from these points. The peaks near strikes/maturities are caused by the infinite Gamma of option payoffs near expiration.[8] As we shall see, these "peaks" in the volatility surface do not affect the continuous dependence of the model values on the input prices: the model value of any contingent claim with a payoff which is continuous except on a set of Wiener measure zero is Lipschitz-continuous with respect to the parameters $C_1, \ldots C_M$.

1.2 Previous approaches to the "implied tree" problem

To our knowledge, the first solution of the implied diffusion problem was proposed by Breeden and Litzenberger (1978), and applied to capital budgeting problems in Banz and Miller (1978). Recently, there have been important contributions by Dupire (1994), Shimko (1993), Rubinstein (1994), Derman and Kani (1994), Barle and Kakici (1995) and Chriss (1996), among others. This "smooth-and-differentiate" approach is based on the observation that a call option price can be written as

$$C(K,T) = \int e^{-rT} \max(S_T - K, 0) p(S_T|S_0) dS_T, \tag{1.6}$$

where $p(S_T|S_0)$ is the conditional probability corresponding to the pricing measure Q associated with the diffusion driving S_t. Differentiating this equation twice with respect to K, we obtain

$$p(K, S_0) = e^{rT} \frac{\partial^2 C(K,T)}{\partial K^2}.$$

This suggests a straightforward way to imply the diffusion driving S_t from option prices. The discrete set of observed option prices is interpolated onto a smooth surface, giving an approximating complete set of prices that can then be numerically differentiated to compute the conditional distribution corresponding to the unknown diffusion.

However, since the price of an option is not uniquely determined in an incomplete market (there is more than one pricing measure), implicit in this approach is the assumption that we can find an "approximating complete market", *before the computation of the transition*

[8] The volatility surface obtained using our method is smooth in the S-variable if the option payoffs are regularized prior to implementing the algorithm.

probabilities. These approaches tend to be unstable since the solution is very sensitive to the smoothness and convexity of the function used in the interpolation.[9]

In Rubinstein (1994), a methodology for constructing an implied binomial tree is described. This method is based on an optimization principle that selects a conditional distribution at some fixed time T that is as close as possible to the distribution corresponding to a standard CRR tree (Cox,Ross and Rubinstein 1979), and that prices a set of options that expire at time T correctly modulo the bid/ask spread.

Rubinstein's approach is revisited in Jackwerth and Rubinstein (1995), where empirical results are discussed and a penalty approach is introduced to smooth the estimated conditional probability function. The fact that Rubinstein's original approach uses only one expiration date has recently been addressed (Jackwerth 1996a; Jackwerth 1996b) . The proposed methodology involves the solution of a large scale optimization problem, with number of variables roughly equal to the number of nodes in the tree.

We mention also the recent paper of Bodhurta and Jermakian (1996) who propose to compute a volatility surface in the form of a perturbation series, where each term in the series is computed by solving a partial differential equation containing source terms determined by the previous term. The coefficients in the partial differential equations are computed as they are required by solving a least-squares problem. This approach effectively solves a series of linear partial differential equations to compute approximate prices and an approximate volatility surface, with the approximation improving as more terms are computed.

In Rubinstein (1994) a least-squares criterion is used to measure the distance between two distributions, but the possible benefits of using other measures, including the relative entropy distance, are discussed. Recent work in the one-period setting has suggested that the relative entropy may be a good choice for such a measure. For example, it is shown in Stutzer (1995) that if we select a distribution that minimizes the relative entropy to a prior subject to pricing constraints, the resulting distribution is maximally unbiased and absolutely continuous with respect to the prior. Relative entropy minimization is also studied in the one-period context in Buchen and Kelly (1996) and Gulko (1995, 1996). The present paper can be seen as an extension of these ideas to the multi-period setting.

1.3 Relationship to the Uncertain Volatility Model

[9] This well-known instability is a consequence of the fact that the problem that we are trying to solve is ill-posed. This is obvious when we compare it to the problem of numerically differentiating a function when we only have discrete noisy observations. At a more fundamental level we note that if we fix T and let K vary in (1.6), we obtain a Volterra integral equation for the transition probabilities. Such equations are known to be ill-posed, and specialized techniques such as regularization, smoothing, filtering, etc., are typically required to solve them (Tikhonov and Arsenin 1977; Banks and Kunish 1989; Banks and Lamm 1985).

In Avellaneda, Levy and Parás (1995) the Uncertain Volatility Model (UVM) was introduced for hedging a position in a portfolio of derivative securities by selecting the worst possible volatility path with respect to this portfolio. This model was combined with a Lagrange multiplier approach in Avellaneda and Parás (1996) in order to minimize the risk of the worst-case hedge by using options as part of the hedge.

There exists a duality between the problem of finding the worst-case volatility path and the problem of implementing a one-sided hedge (that is, one that perfectly protects either a short or a long position). This duality and its game-theoretical implications were studied Samperi (1995), where it was shown that the duality applies even when the derivative claim to be hedged is path-dependent.

The entropy-based approach introduced in this paper can be viewed as an application of the aforementioned framework to a path-dependent "volatility option". Specifically, consider a contingent claim that pays $\int_0^T \eta(\sigma_s^2)ds$ at time T, i.e., pays $\eta(\sigma_s^2)$ for each "day" that the spot volatility is different from the prior.[10] The solution of the stochastic control problem can then be interpreted as the maximum income that an investor with a long position in this claim can earn by hedging his position with the M options. It is worthwhile to point out that this approach can be used to modify the problem by adding other contingent claims to the portfolio to be hedged, thus combining the entropy-minimization idea with the Lagrangian Uncertain Volatility Model (Avellaneda and Parás 1995).

1.4 Outline

In Section 2 we study the notion of Kullback-Leibler relative entropy in the context of diffusions which are mutually singular. This section has the purpose of motivating the constrained stochastic control problem mentioned above.

In Section 3, we present a solution to the stochastic control problem using the Bellman dynamic programming principle, and characterize the calibrated volatility surface in terms of partial differential equations.

In Section 4 we present the basic numerical algorithm, which involves solving simultaneously a system of $M+1$ partial differential equations for the value-function and its gradient with respect to $(\lambda_1, ...\lambda_M)$.

In Section 5, we discuss the qualitative properties of the volatility surface, on the one hand, and present the calculation of "volatility smiles", which consist in interpolation of

[10] This assumes, however, that the spot volatility is observable, which is not the case in practice. Notice also that the payoff is not discounted by the time-value of money, due to the way the pseudo-entropy is derived from the Kullback-Leibler entropy distance (see Section 2). We could also choose to discount the "volatility payoff" at some rate with qualitatively the same results.

the implied volatility data at different maturities. We also analyze the effect of varying the prior, and how this affects the shape of the smile.

In Section 6 we discuss the stability of the method with respect to perturbations in the option prices.

The conclusions are presented in Section 7.

Mathematical proofs which are overly technical or otherwise standard are presented in an Appendix.

2. Minimizing the relative entropy of pricing measures and the constrained stochastic control problem

2.1 Relative entropy of measures in path-space

Given two probability measures P and Q on a common probability space $\{\,\Omega,\,\Sigma\,\}$, the relative entropy, or Kullback-Leibler distance, of Q with respect to P is defined as

$$\mathcal{E}(Q;P) \;=\; \int_{\Omega} ln\left(\frac{dQ}{dP}\right)\,dQ\;, \tag{2.1}$$

where dQ/dP is the Radon-Nikodym derivative of Q with respect to P. $\mathcal{E}(Q;P)$ provides a measure of the relative "information distance" of Q compared to P, where P represents a Bayesian prior distribution. It is well-known that

$(i)\qquad \mathcal{E}(Q;P) \;\geq\; 0\;,$

$(ii)\qquad \mathcal{E}(Q;P) \;=\; 0 \quad\Longleftrightarrow\quad Q \;=\; P\;,$

$(iii)\qquad \mathcal{E}(Q;P) \;=\; \infty \quad$ if Q is not absolutely continuous with respect to P .

Large values of $\mathcal{E}$ correspond to a large information distance (so that Q is very different from the prior P) and $\mathcal{E} \approx 0$ corresponds to low information distance, i.e. proximity to the Bayesian prior P.[11]

We shall study the relative entropy of no-arbitrage pricing measures for derivative securities depending on a single underlying index. Accordingly, consider a pair of probability measures P and Q defined on the set of continuous paths $\Omega = \{ S_\theta \ , \ 0 \leq \theta \leq T \}$ such that

$$\frac{dS_t}{S_t} = \sigma_t^P \, dZ_t^P + \mu_t^P \, dt \ , \quad \text{under } P \tag{2.2a}$$

and

$$\frac{dS_t}{S_t} = \sigma_t^Q \, dZ_t^Q + \mu_t^Q \, dt \ , \quad \text{under } Q \tag{2.2b}$$

in the sense of Itô. Here, $\sigma^\bullet$, $\mu^\bullet$ are assumed to be bounded, progressively measurable processes and $Z^\bullet$ are Brownian motions under the respective probabilities.

The computation of $\mathcal{E}(Q;\, P)$ is straightforward if $\sigma^P = \sigma^Q \equiv \sigma$ with probability 1 under Q. In this case, dQ/dP can be found explicitly using Girsanov's Theorem and we have

$$\mathcal{E}(Q;P) = \frac{1}{2}\,\mathbf{E}^Q \left\{ \int_0^T \left(\frac{\mu_t^Q - \mu_t^P}{\sigma_t} \right)^2 dt \right\} . \tag{2.3}$$

For applications to the calibration of volatility surfaces we should consider situations where the volatilities of the processes in (2.2) are *not* equal with probability 1. In this case the relative entropy is formally equal to $+\infty$, due to the fact that P and Q are mutually singular. To overcome this problem we shall consider discrete-time approximations to these processes and analyze the behavior of the sequence of entropies as the mesh-size tends to zero.

Consider to this end two probability measures P and Q defined on discrete paths

$$S_0,\, S_1,\, ...\, S_N \ ,$$

where N is some integer. The P-probability that such a path occurs can be written as

[11] For background on infomation theory and entropy see Cover and Thomas (1991); Georgescu-Roegen (1971); McLaughlin (1984); Jaynes (1996).

$$\prod_{n=0}^{N-1} \pi_n^P \ ,$$

where π_n^P is the conditional probability given the information set at time n that the price S_{n+1} will occur at date $n+1$. An analogous notation will be used for Q. From (2.1) the relative entropy of Q with respect to P is then given by

$$\begin{aligned}
\mathcal{E}(Q;\,P) &= \sum_{paths} \left(\prod_{n=0}^{N-1} \pi_n^Q \right) \cdot ln \left(\frac{\prod_{n=0}^{N-1} \pi_n^Q}{\prod_{n=0}^{N-1} \pi_n^P} \right) \\
&= \mathbf{E}^Q \left\{ ln \left(\frac{\prod_{n=0}^{N-1} \pi_n^Q}{\prod_{n=0}^{N-1} \pi_n^P} \right) \right\} \\
&= \mathbf{E}^Q \left\{ \sum_{n=0}^{N-1} ln \left(\frac{\pi_n^Q}{\pi_n^P} \right) \right\} \\
&= \mathbf{E}^Q \left\{ \sum_{n=0}^{N-1} \left(\mathbf{E}_n^Q \left[ln \left(\frac{\pi_n^Q}{\pi_n^P} \right) \right] \right) \right\} \ . \qquad (2.4)
\end{aligned}$$

In (2.4), the symbol $\mathbf{E}_n^Q$ represents the conditional expectation operator given the information set at time n. The last equality states that the relative entropy is obtained by summing the *conditional relative entropies* $\mathbf{E}_n^Q \left[ln \left(\frac{\pi_n^Q}{\pi_n^P} \right) \right]$ along each path and averaging with respect to the probability Q.

Let us focus on a special class of approximations to the Itô processes in (2.2) for which the entropy can be computed explicitly as $N \rightarrow \infty$. These processes are based on trinomial trees and are thus well-suited for numerical computation. We assume, specifically, that

$$S_{n+1} = S_n \, H_{n+1} \ , \quad n = 0,\, 1,\, \ldots$$

where

$$H_{n+1} \;=\; \begin{cases} e^{\overline{\sigma}\,\sqrt{dt}} \;, & \text{with probability } P_U \;, \\ \\ 1 \;, & \text{with probability } P_M \;, \\ \\ e^{-\overline{\sigma}\,\sqrt{dt}} \;, & \text{with probability } P_D \;, \end{cases}$$

with transition probabilities given by

$$P_U \;=\; \frac{p}{2}\left(1 \;-\; \frac{\overline{\sigma}\,\sqrt{dt}}{2}\right) \;+\; \frac{\mu\,\sqrt{dt}}{2\overline{\sigma}} \;,$$

$$P_M \;=\; 1 \;-\; p \;,$$

$$P_D \;=\; \frac{p}{2}\left(1 \;+\; \frac{\overline{\sigma}\,\sqrt{dt}}{2}\right) \;-\; \frac{\mu\,\sqrt{dt}}{2\overline{\sigma}} \;. \tag{2.5}$$

Here, $dt \;=\; T/N$ represents the time-step (measured in years). Notice that the logarithm of S_n follows a random walk on the lattice $\left\{\nu\,\overline{\sigma}\,\sqrt{dt}\;,\;\; \nu \text{ integer}\right\}$. In (2.5), the probabilities have been arranged so that the instantaneous mean and variance of $ln\,S_n$ are, respectively, $\mu \;-\; (1/2)\,p\,\overline{\sigma}^2$ and $p\,\overline{\sigma}^2$, consistently with (2.2). Thus, μ and $\overline{\sigma}\,\sqrt{p}$ can be interpreted, respectively, as the carry (interest-rate differential for FX, interest rate minus dividend yield for equities) and the volatility of the index. This model accommodates, by varying the local value of p, processes with variable volatilities in the range $0 \;<\; \sigma_t \;\leq\; \overline{\sigma}$.[12]

The parameters corresponding to the two probabilities P and Q will be denoted by p_0, μ_0 and p, μ respectively. After some computation, we find that[13]

$$\mathbf{E}_n^Q\left[\,ln\left(\frac{\pi_n^Q}{\pi_n^P}\right)\right] \;=\; p\;ln\left(\frac{p}{p_0}\right) \;+\; (1-p)\;ln\left(\frac{1-p}{1-p_0}\right)$$

[12]This last statement is true only for dt small enough so that the probabilities in (2.5) are positive. Notice that this setup produces approximations to diffusion processes (in which the local volatility depends on the price and time-to-maturity) as well as more general random-volatility processes. The latter can be obtained by sampling the volatility from a random distribution.

[13]Notice that the sum of the conditional relative entropies is finite if and only if $p \;=\; p_0$. In this case, the result (2.3) is recovered by replacing the sum of the dt-terms in the right-hand side of (2.6) by an integral. On the other hand, for $p \neq p_0$, the total relative entropy diverges as $dt \;\rightarrow\; 0$.

$$+ \frac{p}{2} \left(\frac{\mu}{p \, \overline{\sigma}} + \frac{\mu_0}{p_0 \, \overline{\sigma}} \right)^2 dt + o(dt) \, , \quad dt \ll 1 \, . \tag{2.6}$$

In the sequel, we shall assume for simplicity that the two processes have identical, constant drift, i.e., $\mu = \mu_0$, that the Bayesian prior P has a constant volatility given by

$$\sigma_0^2 = p_0 \, \overline{\sigma}^2 \, ,$$

and that p varies stochastically under Q. Defining the instantaneous volatility for the Q-process at time $t_n = n \, dt$ by

$$\sigma^2(t_n) = \overline{\sigma}^2 \, p(t_n) \, ,$$

we conclude from (2.6) that the conditional relative entropy at time t_n of Q with respect to P is equal to $\eta\left(\sigma^2(t_n)\right)$ to leading order in dt, where

$$\eta\left(\sigma^2\right) \equiv \frac{\sigma^2}{\overline{\sigma}^2} \, ln \left(\frac{\sigma^2}{\sigma_0^2} \right) + \left(1 - \frac{\sigma^2}{\overline{\sigma}^2} \right) ln \left(\frac{\overline{\sigma}^2 - \sigma^2}{\overline{\sigma}^2 - \sigma_0^2} \right) \, . \tag{2.7}$$

Substituting expression (2.7) into (2.4) and taking into account the estimate of equation (2.6) for the remainder, we conclude that

$$\begin{aligned}
\mathcal{E}(Q;\, P) &= \mathbf{E}^Q \left\{ \sum_{n=0}^{N-1} \left(\, \eta(\sigma^2(t_n)) + O(dt) \, \right) \right\} \\
&= \frac{1}{dt} \, \mathbf{E}^Q \left\{ \sum_{n=0}^{N-1} \eta(\sigma^2(t_n)) \, dt \right\} + O(1) \\
&= \frac{1}{dt} \, \mathbf{E}^Q \left\{ \int_0^T \eta(\sigma^2(t)) \, dt \right\} + O(1) \\
&= N \cdot \frac{1}{T} \, \mathbf{E}^Q \left\{ \int_0^T \eta(\sigma^2(t)) \, dt \right\} + O(1)
\end{aligned}$$

where $T = N \, dt$ and $\mathbf{E}^Q$ represents the expectation operator with respect to the probability distribution of the *continuous-time* process (2.2b). The relevant information-theoretic quantity for $dt \ll 1$ is thus

$$\frac{1}{T}\,\mathbf{E}^Q \left\{ \int_0^T \eta(\sigma^2(t))\,dt \right\} , \tag{2.8}$$

which represents the relative entropy per unit time-step of Q with respect to P.

The notion of entropy per unit time-step is not a property of the Ito processes (2.2), but rather of the pairs of approximating sequences, $(P_N,\, Q_N)$. In fact, the function $\eta(\sigma^2) \approx \mathbf{E}_n^Q \left[ln \left(\frac{\pi_n^Q}{\pi_n^P} \right) \right]$ depends on the discretization used to approximate the pair $(P,\, Q)$. To illustrate the non-uniqueness of η, we consider, for example, a discrete-time approximation of (2.2) in which σ_t^P is constant and σ_t^Q is piecewise constant on time-intervals of length dt. In this case, the single-period distributions are conditionally Gaussian and

$$\eta\left(\sigma^2\right) \;=\; -\frac{1}{2}\left[\, ln\left(\frac{\sigma^2}{\sigma_0^2}\right) \;+\; 1 \;-\; \frac{\sigma^2}{\sigma_0^2}\,\right] . \tag{2.9}$$

Notice that the function η in (2.7) depends on the lattice constant $\overline{\sigma}$. For large values of $\overline{\sigma}$ in (2.7), we have

$$\eta(\sigma^2) \;\approx\; \frac{1}{\overline{\sigma}^2} \left(\, \sigma^2\, ln\left(\frac{\sigma^2}{\sigma_0^2}\right) \;-\; \sigma^2 \;+\; \sigma_0^2 \,\right) .$$

Thus, we may choose to minimize instead the functional (2.8) with

$$\eta(\sigma^2) \;=\; \sigma^2\, ln\left(\frac{\sigma^2}{\sigma_0^2}\right) \;-\; \sigma^2 \;+\; \sigma_0^2 \;. \tag{2.10}$$

2.2 Stochastic control problem

Due to the non-uniqueness of η, it is mathematically convenient to develop a framework for optimization of the functional (2.8) in which $\eta(\sigma^2)$ belongs to a general class of functions which includes (2.7), (2.9) and (2.10) as special cases.

Definition. *A pseudo-entropy (PE) function $\eta(\sigma^2)$ with prior σ_0 is a smooth, real-valued function defined on $0 \,<\, \sigma^2 \,<\, +\infty$, such that*

$(i) \quad 0 \leq \eta(\sigma^2) < \infty ,$

$(ii) \quad \eta(\sigma^2)$ is strictly convex ,

$(iii) \quad \eta(\sigma^2)$ attains the minimum value of zero at $\sigma^2 = \sigma_0^2$.

The reader can easily check that (2.7), (2.9) and (2.10) are PE functions.[14] The simplest PE function with prior σ_0^2 is the quadratic function[15]

$$\eta(\sigma^2) = \frac{1}{2}\left(\sigma^2 - \sigma_0^2\right)^2 \quad , \quad \sigma^2 \geq 0 . \tag{2.11}$$

To model the minimization of the Kullback-Leibler distance in the continuous-time setting, we consider the problem:

Given a pseudo-entropy function η,

$$\text{minimize} \quad \mathbf{E}^Q \left\{ \int_0^T \eta\left(\sigma^2(s)\right) \, ds \right\}$$

$$\text{subject to} \quad \mathbf{E}^Q \left\{ e^{-T_i r} G_i(S_{T_i}) \right\} = C_i , \quad i = 1, \ldots, M$$

among all probability distributions Q *of Itô processes of the form*

$$\frac{dS_t}{S_t} = \sigma_t \, dZ_t + \mu \, dt ,$$

such that σ_t *is a progressively measurable process satisfying* $0 < \sigma_{min} \leq \sigma_t \leq \sigma_{max} < +\infty$.

To avoid degeneracies, we assume that there is a unique option per strike/maturity and that there is at least one strike different from zero.[16]

[14] We note that the function in (2.7) is defined only on the interval $0 < \sigma^2 < \overline{\sigma}^2$. To generate a PE function we can extend it arbitrarily as a convex function for $\sigma^2 > \overline{\sigma}^2$.

[15] This function will be used in numerical computations due to its simplicity. As a rule, the choice of the PE function does not affect qualitatively the results that will follow.

[16] In particular, we do not consider puts and calls with same strike and maturity, since their prices should be exactly related by put-call parity in the absence of arbitrage.

The constraint imposed on σ_t,

$$\sigma_{min} \leq \sigma_t \leq \sigma_{max} \quad , \quad 0 \leq t \leq T \, , \tag{2.12}$$

where σ_{min} and σ_{max} are positive constants, is made for technical reasons. This assumption guarantees that the class of diffusions considered in the control problem is closed with respect to the topology of weak convergence of measures on continuous paths (Billingsley 1968). It is equivalent to the uniform parabolicity of the associated Hamilton-Jacobi-Bellman equation, a desirable feature for achieving stability of standard finite-difference schemes. Specifying *a-priori* bounds on volatility could also be useful in order to incorporate beliefs about extreme volatilities.

We view the the optimization problem as a means to achieving a balance between "subjective beliefs", represented by the prior diffusion

$$\frac{dS_t}{S_t} = \sigma_0 \, dZ_t + \mu \, dt \, ,$$

and the objective information provided by the market prices C_i. Minimization of the relative entropy implies that the pricing measure deviates as little as possible from the prior, while incorporating the observed price information. Thus, entropy minimization corresponds, roughly speaking, to a "minimal" modification of the prior which leads to an arbitrage-free model.[17]

As mentioned in the introduction, the prior plays a significant role in the algorithm. The prior probability determines the behavior of the transition probabilities far away from the mean position (where the information contributed by option prices is "weak" bcause the options have low Gamma). In practice, σ_0 should be chosen so that (a) it is near the implied volatilities corresponding to $C_1, \ldots C_M$, e.g. their geometric or arithmetic mean and (b) it coresponds to the user's expectations about the implied volatility of very low or very high strikes. For instance, to adjust the prior to a market with many expiration dates, one can assume a time-dependent initial prior, $\sigma_0 = \sigma_0(t)$, taking into account the forward-forward volatilities derived from the volatility term-structure. Finally, to incorporate beliefs about the implied volatility at extreme strikes one could consider a prior of the form $\sigma_0 = \sigma_0(S, t)$, with a prescribed behavior for $S \ll 1$ or $S \gg 1$.

3. Solution via dynamic programming

We start with an elementary result from convex duality (Rockafellar, 1970).

[17] While this interpretation is motivated by the calculation of the previous section, it is valid only in an "asymptotic" sense. Here and in the sequel, we refer to a solution of the stochastic control problem as a "minimum-entropy measure" irrespective of the choice of the PE function.

Let η be a PE function with prior σ_0. For $0 \le \sigma_{min} < \sigma_0 < \sigma_{max} \le +\infty$. Define

$$\Phi(X) = \sup_{\sigma_{min}^2 < \sigma^2 < \sigma_{max}^2} \left[\sigma^2 X - \eta(\sigma^2) \right] . \tag{3.1}$$

Lemma 1. *A. If* $\sigma_{max} < +\infty$, *then*

$$(i) \quad \Phi(X) \quad \textit{is convex in } X ,$$

$$(ii) \quad \Phi(0) = 0 ,$$

$$(iii) \quad \Phi'(0) = \sigma_0^2 ,$$

$$(iv) \quad \frac{\Phi(X)}{X} \to \sigma_{max}^2 \quad as \quad X \to +\infty ,$$

$$(v) \quad \frac{\Phi(X)}{X} \to \sigma_{min}^2 \quad as \quad X \to -\infty ,$$

$$(vi) \quad \Phi'(X) = \sigma_{max}^2 \quad for \quad X \ge \eta'(\sigma_{max}^2) ,$$

$$(vii) \quad \Phi'(X) = \sigma_{min}^2 \quad for \quad X \le \eta'(\sigma_{min}^2) .$$

B. If $\sigma_{max} = +\infty$, *and*

$$\lim_{\sigma^2 \to +\infty} \frac{\eta(\sigma^2)}{\sigma^2} = +\infty ,$$

then $\Phi(X)$ *is convex and differentiable for all* X *and (i)-(vii) hold with* $\sigma_{max} = +\infty$.

We shall refer to Φ as the *flux function* associated with the pseudo-entropy η and the bounds σ_{min}, σ_{max}. In the rest of this section, we assume that η, σ_{min} and σ_{max} are fixed and that $0 < \sigma_{min} < \sigma_{max} < \infty$.

There exists a one-to-one correspondence between PE functions and flux functions, in the sense that every flux function satisfying assumptions (i) through (vii) of Lemma 1 corresponds to a PE function. In particular, any monotone-increasing function which interpolates between the values σ_{min}^2 and σ_{max}^2 and takes the intermediate value σ_0^2 at

$X = 0$ can be regarded as the derivative $\Phi'(X)$ of a flux function of a PE function η with prior volatility σ_0.

Proposition 2. *Given a vector of real numbers* $(\lambda_1, \lambda_2, \ldots \lambda_M)$, *let* $W(S, t) = W(S, t; \lambda_1, \lambda_2, \ldots \lambda_M)$ *be the solution of the final-value problem*

$$W_t \;+\; e^{r\,t}\,\Phi\left(\frac{e^{-r\,t}}{2}\,S^2\,W_{SS}\right) \;+\; \mu\,S\,W_S \;-\; r\,W \;=\;$$

$$-\sum_{t\,<\,T_i\,\leq\,T} \lambda_i\,\delta\,(t \,-\, T_i)\;G_i(S)\;, \qquad S \;>\; 0,\; t\;\leq\;T\;, \tag{3.2}$$

with final condition $W(S, T{+}0) \;=\; 0$.[18] *Let* $\mathcal{P}$ *represent the class of probability distributions of admissible Itô processes satisfying (2.12). Then,*

$$W(S,\,t) \;=\; \sup_{Q\,\in\,\mathcal{P}}\; \mathbf{E}_t^Q \left[-\,e^{r\,t}\int_t^T \eta\left(\sigma_s^2\right)\,ds \;+\; \sum_{t\,<\,T_i\,\leq\,T} \lambda_i\,e^{-(T_i-t)}\;G_i\right]\;. \tag{3.3}$$

where $\mathbf{E}_t^Q$ *is the conditional expectation operator with respect to the information set at time* t *and* $S \;=\; S_t$. *Moreover, the supremum in (3.3) is realized by the diffusion process*

$$\frac{dS_t}{S_t} \;=\; \sigma\,(\,S,\,t\,)\;dZ_t \;+\; \mu\,dt\;,$$

with

$$\sigma^2(S,\,t) \;=\; \Phi'\left(\frac{e^{-r\,t}}{2}\,S^2\,W_{SS}\,(S,\,t)\right)\;.$$

The final-value problem (3.2) is well-posed because the partial differential equation is uniformly parabolic. This follows from the properties of Φ listed in Lemma 1. The proof of this Proposition follows the standard procedure for "verification theorems" in Control Theory (Krylov (1980); Fleming and Soner (1992)) It is given in Appendix A.

[18] Subscripts indicate partial derivatives; *e.g.* $W_t \;=\; \partial W/\partial t$, etc. $W(S, T+0)$ represents the value of W for t infinitesimally larger than T. This notation is used to be consistent with the way in which the final conditions corresponding to different option maturities are expressed in (3.2).

Proposition 3. *The function $W(S,\, t;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M)$ defined in Proposition 2 is continuously differentiable and strictly convex in $(\lambda_1,\, \lambda_2,\, \ldots \lambda_M)$.*

For a proof, see also Appendix A. Differentiating equation (3.2) with respect to λ we see that the gradient of $W(S,\, t;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M)$ with respect to the λ variables,

$$W_i \;=\; \frac{\partial W}{\partial \lambda_i},$$

satisfies the partial differential equation

$$W_{i\,t} \;+\; \frac{1}{2}\,\Phi'\left(\frac{e^{-rt}}{2} S^2\, W_{SS}\right)\, S^2\, W_{i\,SS} \;+\; \mu\, S\, W_{i\,S} \;-\; r\, W_i \;=\; -\,\delta(t - T_i)\, G_i\;, \qquad (3.4)$$

for $1 \;\le\; i \;\le\; M$, with $W_i(S,\, T+0) \;=\; 0$. These equations can be interpreted as pricing equations for the M input options using the diffusion with volatility $\sigma^2(S,\, t) \;=\; \Phi'\left(\frac{e^{-rt}\, S^2\, W_{SS}(S,\, t)}{2}\right)$. In particular, the model will be calibrated if

$$W_i(S,\, 0) \;=\; C_i \quad, \qquad \text{or} \qquad \frac{\partial W(S,\, 0;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M)}{\partial \lambda_i} \;=\; C_i\;.$$

This shows that calibration is equivalent to minimizing the function $W(S,\, 0;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M) - \sum \lambda_i\, C_i$. The next proposition formalizes this and shows that this choice of volatility solves the stochastic control problem.

Proposition 4. *Define*

$$V(S,\, t;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M\,) \;=\; W(S,\, t;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M\,) \;-\; \sum_{i=0}^{M} \lambda_i\, C_i\;.$$

Suppose that, for fixed S, $V(S,\, 0;\, \lambda_1,\, \lambda_2,\, \ldots \lambda_M\,)$ attains a global minimum at the point $(\lambda_1^*,\, \lambda_2^*,\, \ldots \lambda_M^*\,)$ in $\mathbf{R}^M$. Then, the class of probability measures satisfying the price constraints and the volatility bounds (2.12) is non-empty and the stochastic control problem problem admits a unique solution. The solution corresponds to the diffusion process with volatility

$$\sigma(S,\, t) \;=\; \sqrt{\Phi'\left(\frac{e^{-r\,t}}{2}\, S^2\, W_{SS}\right)}\;, \qquad 0 \;\le\; t \;\le\; T\;,$$

where W is the solution of the final-value problem (3.2) with $\lambda_i \; = \; \lambda_i^*$, $i \; = \; 1, \; ... \, M$.

Proof. To establish (i), observe that

$$V\,(\,S,\,0;\;\lambda_1,\lambda_2,\,...\,\lambda_M\,) \;=\; \sup_{Q\,\in\mathcal{P}} \left(\; a(Q) \;+\; \sum_{i=1}^{M} \lambda_i\, b_i(Q) \;\right) \;, \tag{3.5}$$

where

$$a(Q) \;=\; \mathbf{E}^Q \left\{ \;-\; \int_0^T \;\eta(\,\sigma_s^2\,)\,ds \;\right\}$$

and

$$b_i(Q) \;=\; \mathbf{E}^Q \left\{\; e^{-r\,T_i}\, G_i(S_{T_i}) \;\right\} \;-\; C_i \quad , \quad i \;=\; 1,\,2\,...\,M \;.$$

Suppose the function attains a global minimum at some M-tuple $(\lambda_1^*,\, ...\lambda_M^*\,)$ and let Q^* denote the unique measure that realizes the sup in (3.5) for these λ-values. (The measure Q^* is unique, by Proposition 2.) The linear function

$$a(Q^*) \;+\; \sum_{i=1}^{M} \lambda_i\, b_i(Q^*)$$

can be viewed as the graph of a supporting hyper-plane to the graph of V passing through the minimum. In particular, the smoothness of V (a consequence of Proposition 2) implies that this hyper-plane is tangent to the graph of V and thus that

$$\left(\frac{\partial V}{\partial \lambda_i}\right)_{\lambda \;=\; \lambda^*} \;=\; b_i(Q^*) \;=\; \mathbf{E}^{Q^*} \left\{\; e^{-r\,T_i}\, G_i(S_{T_i}) \;\right\} \;-\; C_i \;=\; 0$$

for $i \;=\; 1\,...\,M$. The subset of measures of type $\mathcal{P}$ which satisfy the price constraints is therefore non-empty: it contains at least the element Q^*.

Suppose now that Q' is another measure in the class $\mathcal{P}$ such that

$$\mathbf{E}^{Q'} \left\{\; e^{-r\,T_i}\, G_i(S_{T_i}) \;\right\} \;=\; C_i \;, \quad i = 1\,...\,M \;.$$

Then, $b_i(Q') \;=\; 0$, so

$$a(Q') \;=\; a(Q') \;+\; \sum_{i=1}^{M} \lambda_i^*\, b_i(Q')$$

$$\leq\; \sup_{Q\,\in\mathcal{P}} \left(a(Q) \;+\; \sum_{i=1}^{M} \lambda_i^*\, b_i(Q) \right) \;,$$

$$=\; a(Q^*) \;.$$

This establishes that Q^* has the smallest relative entropy among all measures of type $\mathcal{P}$ satisfying the price constraints.

4. Numerical Implementation

The numerical solution consists in computing the function $V(S,\,0;\,\lambda_1,\,...\,\lambda_M)$ and searching for its minimum in λ-space. For this purpose, we consider a system of PDEs for the evaluation of this function and its derivatives,

$$V_i(S,\,0;\,\lambda_1,\,...\,\lambda_M) \;=\; W_i(S,\,0;\,\lambda_1,\,...\,\lambda_M) \;-\; C_i \;,\;\; i \;\leq\; i \;\leq\; M \;,$$

namely,

$$V_t \;+\; e^{r\,t}\Phi\left(\frac{e^{-r\,t}}{2} S^2\, V_{SS}\right) +\; \mu\, S\, V_S \;-\; r\, V \;=$$

$$-\sum_{t\,<\,T_i}^{M} \lambda_i \;\left(G_i(S_{T_i}) \;-\; e^{r\,T_i}\, C_i\right)\, \delta(t\;-\;T_i) \;, \qquad (4.1)$$

$$V_{i\,t} \;+\; \frac{1}{2}\,\Phi'\left(\frac{e^{-r\,t}}{2} S^2\, V_{SS}\right)\;\; S^2\, V_{i\,SS} +\; \mu\, S\, V_{i\,S} \;-\; r\, V_i \;=$$

$$- \sum_{t < T_i}^{M} \left(G_i(S_{T_i}) \; - \; e^{r\,T_i}\, C_i \right) \, \delta(t \; - \; T_i) \;, \tag{4.2}$$

for $1 \;\leq\; i \;\leq\; M$.

Concretely, the algorithm for finding the minimum of $V(S,\,0;\,\lambda_1,\,...\,\lambda_M)$ consists in

- rolling back the values of the vector $(V,\,V_1,\,...\,V_M\,)$ to the date $t = 0$,
- updating the estimate of $(\lambda_1,\,...\,\lambda_M)$ using the computed value of the gradient with a gradient-based optimization subroutine,
- repeating the above steps until the minimum is found.

Our numerical method for solving (4.1)-(4.2), uses a finite-difference scheme (trinomial tree) presented in Section 2.1, with the risk-neutral probabilities in (2.5). We implemented, for simplicity, the quadratic pseudo-entropy function in (2.11).

The corresponding flux function is

$$\Phi(X) \;=\; \begin{cases} \frac{1}{2},X^2 \;+\; \sigma_0^2\,X \;\;, & \sigma_{min}^2 \;-\; \sigma_0^2 \;<X\;<\; \sigma_{max}^2 \;-\; \sigma_0^2 \;, \\[2ex] \sigma_{min}^2\,X \;-\frac{1}{2}\left(\sigma_{min}^2 -\; \sigma_0^2\right)^2 \;, & X \;\leq\; \sigma_{min}^2 \;-\; \sigma_0^2 \;, \\[2ex] \sigma_{max}^2\,X \;-\frac{1}{2}\left(\sigma_{max}^2 -\; \sigma_0^2\right)^2 & X \;\geq\; \sigma_{max}^2 \;-\; \sigma_0^2 \;, \end{cases}$$

The derivative of Φ varies linearly between σ_{min}^2 and σ_{max}^2. It is given by

$$\Phi'(X) \;=\; \begin{cases} X \;+\; \sigma_0^2 \;\;, & \sigma_{min}^2 \;-\; \sigma_0^2 \;\leq\; X \;\leq\; \sigma_{max}^2 \;-\; \sigma_0^2 \;, \\[2ex] \sigma_{min}^2 \;\;, & X \leq\; \sigma_{min}^2 \;-\; \sigma_0^2 \;, \\[2ex] \sigma_{max}^2 \;\;, & X \;\geq\; \sigma_{max}^2 \;-\; \sigma_0^2 \;. \end{cases}$$

As a numerical approximation for the "dollar Gamma" $\frac{1}{2}S^2\,V_{SS}$ in the lattice, we take

$$\left(\frac{1}{2} S^2 \, V_{SS} \right)_n^j \quad \longleftrightarrow$$

$$\frac{1}{\overline{\sigma}^2\,dt} \left[\left(1 \;-\; \frac{\overline{\sigma}\,\sqrt{dt}}{2} \right) \cdot V_n^{j+1} \;+\; \left(1 \;+\; \frac{\overline{\sigma}\,\sqrt{dt}}{2} \right) \cdot V_n^{j-1} \;-\; 2\,V_n^j \right] . \tag{4.3}$$

The partial differential equations are approximated by local "roll-backs" using the probabilities (2.5) with the appropriate choice for the parameter p at each node, dictated by the value of (4.3). The "local volatility" in the trinomial tree is

$$\left(\sigma_n^j\right)^2 = p_n^j \, \overline{\sigma}^2 ,$$

so we take

$$p_n^j = \frac{1}{\overline{\sigma}^2} \left[\frac{\Phi\left(e^{-r\,t}\left(\frac{1}{2}S^2\, V_{SS}\right)_n^j\right)}{e^{-r\,t}\left(\frac{1}{2}\, S^2\, V_{SS}\right)_n^j} \right]$$

in equation (4.1) and

$$p_n^j = \frac{1}{\overline{\sigma}^2}\, \Phi'\left(e^{-r\,t}\left(\frac{1}{2}\, S^2\, V_{i,SS}\right)_n^j\right)$$

for equation (4.2).

The scheme implemented for this study was explicit Euler with trimming of the tails after 3.5 standard deviations.[19] For the numerical optimization, we used the BFGS algorithm(Byrd *et al* (1994); Byrd *et al* (1996); Zhu *et al* (1994)).

5. The volatility surface

5.1 Spot volatility

We study in more detail the spot volatility surface computed by this algorithm. To simplify the analysis, we perform a change of variables that eliminates μ and r from the right-hand side of the PDE (38), namely:

$$\tilde{V} = e^{-r\,t}\, V\ , \qquad \tilde{S} = e^{-\mu\, t}\, S\ .$$

With these new variables[20], equation (4.1) becomes

$$\tilde{V}_t + \frac{1}{2}\,\Phi\left(\frac{\tilde{S}^2\,\tilde{V}_{SS}}{2}\right) = -\sum_{T_i\ <\ t \le T}^{M} \lambda_i\,\left[e^{-r\,T_i}\, G_i(\tilde{S}\, e^{\mu\, T_i}) - C_i\right]\,\delta(t - T_i)\ , \qquad (5.1)$$

[19] See Parás (1995) for a proof of consistency of the scheme.

[20] The new variables correspond to the value of assets measured in dollars at time $t = 0$.

with $\tilde{V}(S,\, T+0) \;=\; 0$.

Differentiating this equation twice with respect to $\tilde{S}$ and multiplying both sides by $\frac{\tilde{S}^2}{2}$ we obtain an evolution equation for the "dollar-Gamma" of the value function

$$\tilde{\Gamma} \;=\; \frac{\tilde{S}^2\,\tilde{V}_{\tilde{S}\tilde{S}}}{2} \;=\; \frac{e^{-r\,t}\,S^2\,V_{SS}}{2}\,.$$

Dropping the tildes to simplify notation, the equation thus obtained is

$$\Gamma_t \;+\; \frac{S^2}{2}\,(\;\Phi(\Gamma)\;)_{SS} \;=$$

$$-\sum_{T_i\;<\;t\leq T}^{M} \lambda_i\, e^{-(r-\mu)\,T_i}\,\delta\left(S \;-\; e^{-\mu\,T_i}\,K_i\,\right)\delta(t\;-\;T_i)\;, \tag{5.2}$$

or

$$\Gamma_t \;+\; \frac{S^2}{2}\,(\;\Phi'(\Gamma)\,\Gamma_S\;)_S \;=$$

$$-\sum_{T_i\;<\;t\leq T}^{M} \lambda_i\, e^{-(r-\mu)\,T_i}\,\delta\left(S \;-\; e^{-\mu\,T_i}\,K_i\,\right)\delta(t\;-\;T_i)\;. \tag{5.3}$$

The latter equation clarifies the nature of the volatility surface

$$\sigma^2 \;=\; \Phi'\,(\Gamma)\;\;. \tag{5.4}$$

For instance, if the option prices C_i are exactly the Black-Scholes prices with volatility σ_0, the solution of the stochastic control problem has $\lambda_i^* \;=\; 0$ for all i and $\Gamma \;=\; 0$, consistently with the fact that $\sigma^2(S,t) \;=\; \sigma_0^2$ is the minimum-entropy solution. (In this case no information is added by considering option prices.) On the other hand, if one or more option prices are inconsistent with the prior, the Lagrange multipliers are not all zero. Each non-zero λ_i^*, gives rise to a Dirac source in (5.2)-(5.3). The resulting Γ profile is initially singular (it is similar to the Gamma of an option portfolio) and diffuses progressively into the $(S,\, t)$-plane as a smooth function. Instantaneous smoothing of Γ is guaranteed by the bounds on the volatility Φ' which follow from (2.12) (cf. Lemma 1). Using equation (5.4), we find that, immediately before time T_i and near the strike, σ^2 is equal to σ_{min} or σ_{max}, according to the sign of λ_i^*. As $T_i \;-\; t$ increases, the surface becomes smoother and the constraint $\sigma_{min} \;\leq\; \sigma_t \;\leq\; \sigma_{max}$ is non-binding. Generically, each point $(K_i,\, T_i)$ gives rise to a disturbance of the volatility surface, which looks like a "ridge" ($\lambda_i^* \;>\; 0$) or a "trough" ($\lambda_i^* \;<\; 0$). To complete the picture, note that the disturbances "interact" with each other due to the nonlinearity of the equation. The overall topography of the surface is determined by the relative strengths of the Lagrange multipliers λ_1^*, ... λ_M^*.[21]

5.2 Implied volatility: interpolating between traded strikes

The main application of the implied volatility surface is to calculate the fair values of derivative securities which are not among the M input options. An interesting diagnostic for our algorithm consists in analyzing the implied volatility profiles that can be generated after calibrating the model to a finite number of option prices. There are two features of interest here: the shape of the curve between strikes (interpolation) and the shape of the curve for strikes which are smaller or larger than the ones used for calibration (extrapolation).

A first set of numerical experiments was done using the Dollar/Mark dataset of Appendix B; cf. Figures 1 and 2. At each of the standard maturities, ranging from 30 days to 270 days, we have 5 traded strikes. After calibrating to the mid-market prices of these options using the parabolic PE function with prior $\sigma_0 = 0.141$ (a rough average of the implied volatilities of traded options), we computed option prices for a sequence of strikes at each expiration date using a fine mesh. We then computed the corresponding implied volatilities and generated an "implied smile" for each standard maturity.

The curves are shown in Figure 3.

Notice that the shapes are influenced by the relation between the implied volatilities and the prior. This market corresponds to an "inverted" volatility term-structure, with near-term options trading at more than 14% or higher and 270-day options trading at approximately 13% volatility.

Given our choice of prior (arbitrarily chosen), σ_0 is lower than the volatilities of traded options with short-maturities and higher than the implied volatilities of traded options with long maturities. The minimum relative-entropy criterion tends to "pull" the implied volatility curve towards the prior. The "pull-to-prior" effect can be seen in the way the curve interpolates between strikes. For low priors, the interpolation tends to be a convex curve while for high priors the interpolated curve tends to be concave.

The "wings" of the implied volatility curves are lower than the (extreme) 20-delta volatilities for short-term options, higher than the 20-delta volatilities for long-term options and are practically horizontal for the 90-day puts and 60-day calls, that have volatilities approximately equal to the prior. In all cases, the extreme values of the volatility tend to the prior volatility, as we expect.

[21] In numerical calculations, point sources corresponding to small values of λ^* may not always be observable, due to the discrete approximation of the Delta functions. Thus, weak point sources may become "masked" by the Γ produced by other options with larger λ^*.

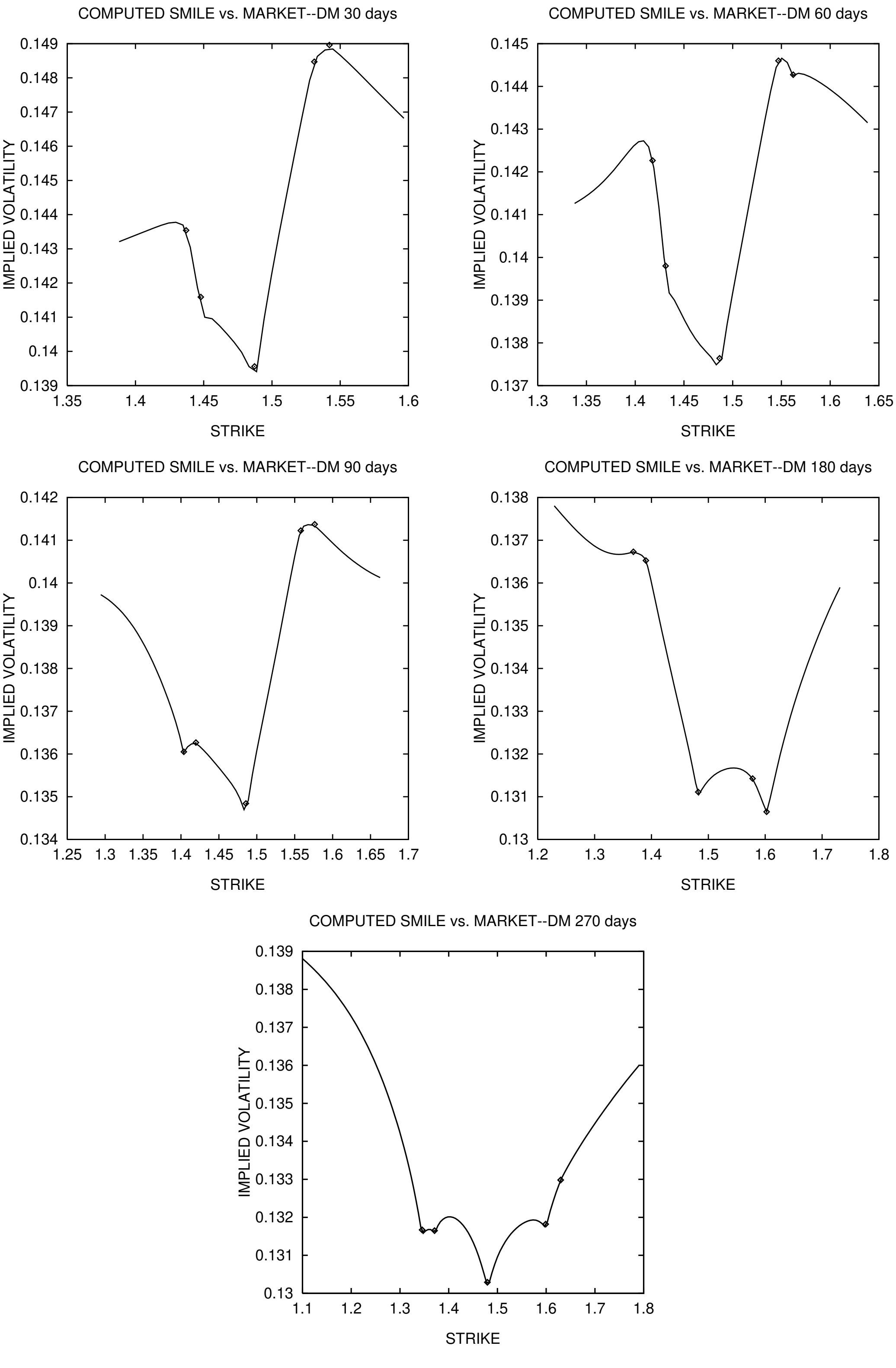

FIGURE 3. Implied volatility curves at different expiration dates computed using a constant prior of 0.141. The data is given in Appendix B.

This calculation show that, in practice, it may be necessary to consider prior volatilities that depend on both S and t. A more conventional form of the smile could then be achieved by choosing σ_0 using the term-structure of volatility of at-the-money-forward options for S between traded strikes and a higher prior to extrapolate beyond traded strikes.

To investigate in more detail the effect of the prior on the interpolation between traded strikes, we considered a hypothetical market with three traded options, expiring in 30 days, with strikes equal to 100, 95 and 105 percent of the spot price. We assumed that the implied volatilities of the options were 14%, 15% and 16%, respectively and that $\mu = r = 0$. We calibrated four different volatility surfaces for this dataset, using priors of 11%, 13%, 14% and 17%. The results are displayed in Figure 4. These calculations confirms our previous conclusions on the sensitivity of the implied volatility curve to the prior.

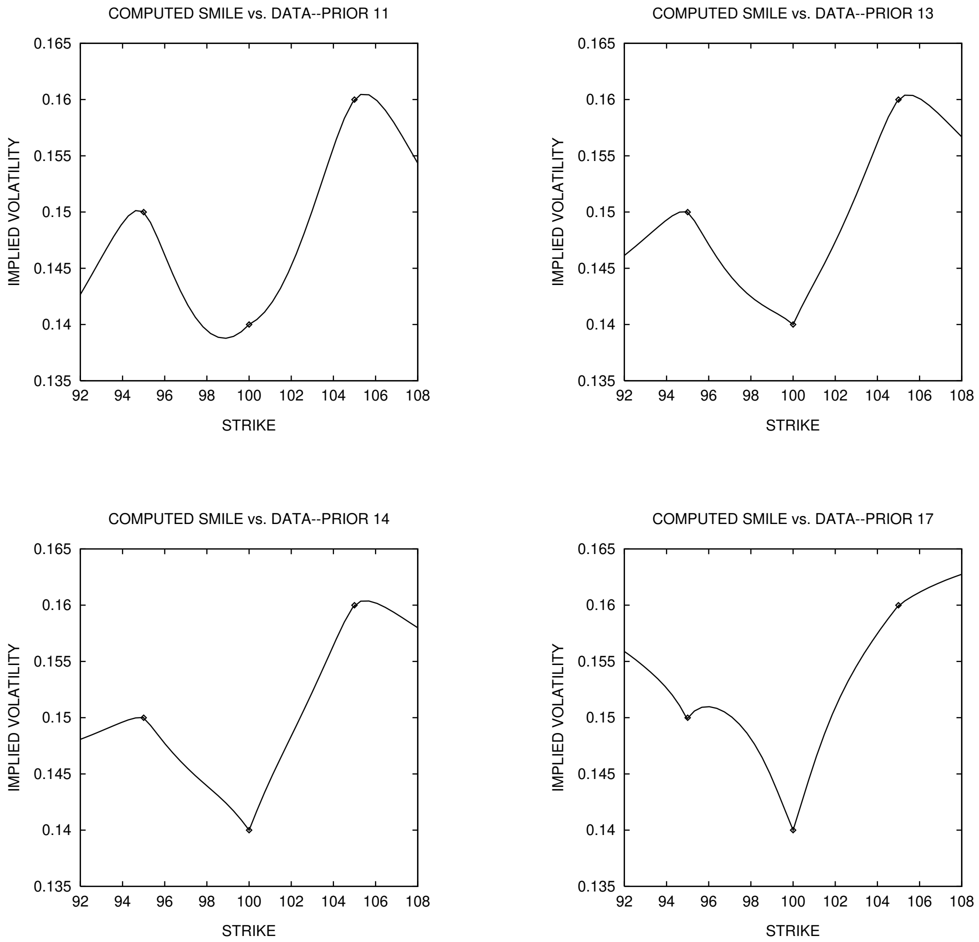

FIGURE 4. Effect of varying the prior volatility on the interpolated implied volatility curve (smile). The data consists of 3 options with maturity 30 day and volatilities 14%(strike=100), 15%(strike=95) and 16%(strike=105). Interest rates were taken to be zero.

6. Convex duality, Lagrange multipliers and stability analysis

We may visualize the solution of the optimality problem by considering the function

$$W\,(S,\,0;\,\lambda_1,\,...\,\lambda_M)\ ,$$

defined in equation (3.2). This function depends on $r,\,\mu,\,K_i,\,T_i,\,i\;=\;1,\,...\,M$, on the pseudo-entropy function η and on the volatility bounds σ_{min} and σ_{max}. We have established that W is smooth and strictly convex in $(\lambda_1,\,...\,\lambda_M)$. Solving the optimization problem corresponds therefore to finding, for a given a price vector $(\,C_1,\,...\,C_M)$, the quantity

$$\begin{aligned} U\,(\,C_1,\,...\,C_M\,) \;&\equiv\; \inf_{\lambda_1,\,...\,\lambda_M} \;\; V\,(S,\,0;\,\lambda_1,\,...\,\lambda_M\,) \\ &=\; \inf_{\lambda_1,\,...\,\lambda_M} \left[\, W\,(S,\,0;\,\lambda_1,\,...\,\lambda_M\,) \;-\; \sum_{i=1}^{M} \lambda_i\,C_i \,\right]\,. \end{aligned} \tag{6.1}$$

and the Lagrange multipliers. The function $U\,(\,C_1,\,...\,C_M\,)$ represents the "maximum entropy per lattice site" of measures in the class $\mathcal{P}$ (i.e. Ito processes with drift μ and volatility satisfying the *a priori* volatility bounds) which match market prices. It is the dual of W, in the sense of convex duality.[22]

Geometrically, $U\,(\,C_1,\,...\,C_M\,)$ corresponds to the largest value of a for which the hyperplane in $\mathbf{R}^{M+1}$

$$h_a\,(\lambda_1,\,...\,\lambda_M) \;=\; \sum_{i=1}^{M} \lambda_i\,C_i \;+\; a$$

satisfies $h_a\,(\lambda_1,\,...\,\lambda_M) \;\le\; W\,(S,\,0;\,\lambda_1,\,...\,\lambda_M)$ for all $(\lambda_1,\,...\,\lambda_M)$. Notice that these hyper-planes are normal to the direction

$$(\,C_1,\,...\,C_M,\,-1)\ .$$

Therefore, *the stochastic control problem admits a solution if and only if the price vector* $(\,C_1,\,...\,C_M,\,-1)$ *belongs to the cone of normal directions to the graph of* W. If $(C_1,\,...\,C_M)$

[22]Strictly speaking, $-U$ is the Legendre dual of W. The functions η and Φ are in a similar correspondence, if we redefine η to be $+\infty$ for σ^2 outside the interval $[\,\sigma_{min},\,\sigma_{max}\,]$.

satisfies this condition, the Lagrange multipliers correspond to points of contact between of the optimal hyper-plane h_a with the graph of W.

It is noteworthy that the cone of normal directions to the graph of W, and hence the domain of U, is *independent of the choice of entropy.* In fact, it coincides with the cone generated by the vectors

$$\left\{\, \mathbf{E}^Q\, \left[e^{-r\,T_1}\, G_1(S_{T_1}) \right],\ \ldots,\ \mathbf{E}^Q\, \left[e^{-r\,T_M}\, G_1(S_{T_M}) \right],\ -1 \,\right\} \tag{6.2}$$

as Q varies in the class $\mathcal{P}$.[23]

The next proposition is an immediate consequence of the strict convexity of W and convex duality.

Proposition 5. *$U(C_1, \ldots C_M)$ is a concave function of class $C^{1,1}$ in the interior of its domain of definition. The Lagrange multipliers $\lambda_1^*, \ldots \lambda_M^*$ are differentiable functions of the price vector and satisfy*

$$\lambda_i^* \;=\; -\,\frac{\partial U}{\partial C_i} \quad , \qquad \frac{\partial \lambda_i^*}{\partial C_j} \;=\; -\,\frac{\partial^2 U}{\partial C_i\, \partial C_j}\;.$$

Moreover,

$$-\,\frac{\partial^2 U}{\partial C_i\, \partial C_j} \;\; \textit{is the inverse matrix of} \;\; \frac{\partial^2 W}{\partial \lambda_i\, \partial \lambda_j}\;.$$

Thus, if $(C_1, \ldots C_M)$ varies in a compact subset of the domain of U, the sensitivities $\frac{\partial \lambda_i^*}{\partial C_j}$ remain uniformly bounded. We conclude that the functions $W(S,\,t)$ and $W_S(S,\,t)$ are Lipschitz continuous functions of the C_i, uniformly in (S, t). The same is true for the second derivative W_{SS} in any closed region of the $(S,\,t)$-plane which excludes the points $(K_i,\,T_i)$, $i \,=\, 1, \ldots M$. At these points, the second derivative of W is singular, because $G_{i,SS} \;=\; \delta(S - K_i)$ and hence $W_{SS}(S,\,T_i) \;=\; \lambda_i^*\, \delta(S - K_i)$. A discontinuity of $W_{SS}(K_i,\,T_i)$ with respect to $(C_1, \ldots C_M)$ will occur when the Lagrange multiplier λ_i^* crosses zero and W_{SS} changes sign.[24] In particular, the volatility surface

$$\sigma\,(S,\,t) \;=\; \sqrt{\Phi'\left(\frac{e^{-r\,t}\, S^2\, W_{SS}(S,\,t)}{2} \right)}$$

[23] The reason for this is that the latter cone is the tangent cone at infinity to the cone of normals to W.

[24] We note, however, that W_{SS} is Lipschitz continuous in C_i as a signed measure.

is uniformly Lipschitz-continuous as a function of $(C_1, .. C_M)$ for (S, t) bounded away from the points (K_i, T_i) , $i = 1, ... M$. Note, however, that the prices of contingent claims obtained with this model are continuous in $(C_1, .. C_M)$, since W depends smoothly on the Lagrange multipliers and hence on the price vector. The algorithm is therefore stable with respect to perturbations of the price vector.

The stability of the algorithm deteriorates, however, as the price vector approaches the boundary of the domain of definition of U, due to the fact that the Lagrange multipliers increase indefinitely and U tends to $-\infty$ as $(C_1, .. C_M, -1)$ approaches the boundary of the cone (6.2). To increase the stability of numerical computations in these cases, the volatility band should be widened until the Lagrange multipliers are of order 1.

7. Conclusions

The calibration of a diffusion model to a set of option prices can be cast as a minimax problem which corresponds to the minimization of the relative entropy distance between the surface that we wish to find and a Bayesian prior distribution.

The minimax problem can be solved by dynamic programming combined with the minimization of a function of M variables, where M is the number of prices that we seek to match. The evaluation of the function that we wish to minimize and of its gradient is done by solving a system of $M + 1$ partial differential equations on a trinomial tree.

The resulting volatility surfaces are essentially the minimal perturbations of the Bayesian prior that match all option prices. Accordingly, the method allows for constructing a surface that takes into account not only option prices but also the user's expectations about volatility (via the prior). Qualitatively, the surface consists of ridges or troughs superimposed on the prior surface, which are sharp near the strike/expiration points (K_i, T_i) and diffuse smoothly away from these points. Roughly speaking, the shapes of the distortions are close to the shape of the Gamma-surface of an option.

We have shown that the prices of contingent claims generated by the model vary continuously with the input option prices $(C_1, ... C_M)$. The stability and height of the volatility surface at the strike/maturity points is controlled by the bounds σ_{min} and σ_{max}.

Numerical calculations show that the algorithm can be used to interpolate between the implied volatilities of traded options. The curves obtained in this fashion depend, however, on the choice of prior distribution. In particular, for extrapolation beyond traded strikes, prior volatilities that take into account subjective expectations about volatilities conditional upon extreme market moves should be used. These and other qualitative features of the algorithm will be studied in future publications.

References

Avellaneda, M., A. Levy, and A. Parás (1995). Pricing and hedging derivative securities in markets with uncertain volatilities. *Applied Mathematical Finance 2*, 73–88.

Avellaneda, M. and A. Parás (1996). Managing the volatility risk of portfolios of derivative securities: The Lagrangian uncertain volatility model. *Applied Mathematical Finance 3*, 21-52

Banks, H. and P. D. Lamm (1985). Estimation of variable coefficients in parabolic distributed systems. *IEEE Transactions on Automatic Control 30*(4), 386–398.

Banks, H. T. and K. Kunisch (1989). *Estimation Techniques for Distributed Parameter Systems.* Boston: Birkhäuser.

Banz, R. W. and M. H. Miller (1978). Prices for state-contingent claims: Some estimates and applications. *Journal of Business 51*(4), 653–672.

Barle, S. and N. Cakici (1995, October). Growing a smiling tree. *Risk Magazine 8*(10), 76–81.

Billingsley, P. (1968). *Convergence of Probability Measures*, New York, John Wiley & Sons.

Bodurtha, Jr., J. N. and M. Jermakyan (1996, October). Non-parametric estimation of an implied volatility surface. Georgetown University working paper.

Breeden, D. T. and R. H. Litzenberger (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business 51*(4), 621–651.

Buchen, P. W. and M. Kelly (1996, March). The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis 31*(1), 143–159.

Byrd, R. H., P. Lu, J. Nocedal and C. Zhu (1994, May). A Limited Memory Algorithm for Bound Constrained Optimization, Northewstern University, Department of Electrical Engineering NAM-08, Evanston Ill.

Byrd, R. H., J. Nocedal and R. B. Schnabel (1996, January). Representation of Quasi-Newton Matrices and their use in Limited Memory Models, Northwestern University, Department of Electrical Engineering NAM-03, Evanston Ill.

Chriss, N. (1996, July). Transatlantic trees. *Risk Magazine 9*(7).

Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory.* New York: John Wiley & Sons.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979). Option pricing: A simplified approach. *Journal of Financial Economics 7*, 229–263.

Derman, E. and I. Kani (1994, February). Riding on a smile. *Risk Magazine 7*(2).

Dupire, B. (1994, January). Pricing with a smile. *Risk Magazine 7*(1).

Fleming, W. H. and M. Soner (1992). *Controlled Markov Processes and Viscosity Solutions*, Springer-Verlag, New York.

Friedman, A. (1964). *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, N.J.

Georgescu-Roegen, N. (1971). *The Entropy Law and the Economic Process.* Cambridge, Massachusetts: Harvard University Press.

Gulko, L. (1995, October). The entropy theory of bond option pricing. Yale University Working Paper.

Gulko, L. (1996, May). The entropy theory of stock option pricing. Yale University Working Paper.

Jackwerth, J. C. (1996a, August). Generalized binomial trees. Berkeley Working Paper.

Jackwerth, J. C. (1996b, August). Implied binomial trees: Generalizations and empirical tests. Berkeley Working Paper.

Jackwerth, J. C. and M. Rubinstein (1996). Recovering probability distributions from contemporaneous security prices. *Journal of Finance*, 51:5 1611-1631, December.

Jaynes, E. T. (1996, March). Probability theory: The logic of science. Unpublished Manuscript, Washington University, St. Louis, MO.

Krylov, N. V. (1980). *Controlled Diffusion Processes*, Springer-Verlag, New York.

McLaughlin, D. W. (Ed.) (1984, April). *Inverse Problems*, Providence, Rhode Island. SIAM and AMS: American Mathematical Society. Volume 14.

Parás, A. (1995). *Non-linear partial differential equations in finance: a study of volatility risk and transaction costs*, Ph. D. Thesis, New York University.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton, New Jersey: Princeton University Press.

Rubinstein, M. (1994, July). Implied binomial trees. *The Journal of Finance 69*(3), 771–818.

Samperi, D. J. (1995). Implied trees in incomplete markets. Courant Institute Research Report, New York University.

Shimko, D., (1993). Bounds on Probability, *Risk magazine* October, 59-66.

Stutzer, M. (1995). A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics 68*, 367–397.

Tikhonov, A. N. and V. Y. Arsenin (1977). *Solutions of Ill-Posed Problems.* New York: Wiley.

Zhu, C., R. H. Boyd, P. Lu, and J. Nocedal, (1994, December). L-BFGS-B: FORTRAN Subroutines for Large-Scale Bound Constrained Optimization, Northwestern University, Department of Electrical Engineering Evanston Ill.

Appendix A

A.1 Proof of Proposition 2

Consider the Itô process

$$\frac{dS_t}{S_t} = \mu dt + \sigma_t dZ_t \ ,$$

where σ_t is an adapted random process such that

$$\sigma^2_{min} \ \leq \ \sigma^2 \ \leq \ \sigma^2_{max} \ .$$

By Itô's Lemma, we have

$$d \left(e^{-r\,t} \, W \left(S_t , \, t \right) \right) \ =$$

$$e^{-r\,t} \cdot \left\{ W_S \left(S_t , \, t \right) \, dS_t \ + \ \left[\ W_t \ + \ \frac{1}{2} \, \sigma^2_t \, S^2_t \, W_{SS} \left(S_t , \, t \right) \ - \ r \, W (S_t , \, t) \right] \, dt \ \right\} \ . \tag{A.1}$$

From the inequality

$$X \, \sigma^2 \ \leq \ \Phi \left(X \right) \ + \ \ \eta (\sigma^2) \ \ , \tag{A.2}$$

which follows from the definition of Φ, we have

$$\begin{aligned} \frac{1}{2} \, S^2 \, W_{SS} \cdot \sigma^2 \ \ &= \ e^{r\,t} \cdot \left(\frac{e^{-r\,t}}{2} \, S^2 \, W_{SS} \cdot \sigma^2 \right) \\ &\leq \ e^{r\,t} \cdot \Phi \left(\frac{e^{-r\,t}}{2} \, S^2 \, W_{SS} \right) \ + \ e^{r\,t} \, \eta (\sigma^2_t) \ \ . \end{aligned} \tag{A.3}$$

Substituting this inequality into (A.1) and rearranging terms, we obtain

$$
\begin{aligned}
& d\left(e^{-r\,t}\, W(S_t,\, t) \right) \\
&= e^{-r\,t} W_S\, S_t\, \sigma_t^2\, dZ_t \;+\; e^{-r\,t} \left\{ W_t \;+\; \frac{1}{2}\, \sigma_t^2\, S_t^2\, W_{SS} \;+\; \mu\, S_t \;-\; r\, W \right\} dt \\
&\le e^{-r\,t} W_S\, S_t\, \sigma_t^2\, dZ_t \\
&\quad + e^{-r\,t} \left\{ W_t \;+\; e^{r\,t}\, \Phi\left(\frac{e^{-r\,t}}{2}\, S_t^2\, W_{SS} \right) \;+\; \mu\, S_t \;-\; r\, W \right\} dt \;+\; \eta\left(\sigma_t^2 \right)\, dt \\
&= e^{-r\,t} W_S\, S_t\, \sigma_t^2\, dZ_t \;-\; \sum_{t\,<\,T_i\,\le\,T} \lambda_i\, e^{-r\,t}\, \delta\,(t \,-\, T_i)\; G_i(S_t)\, dt \;+\; \eta\left(\sigma_t^2 \right)\, dt\ ,
\end{aligned}
\tag{A.4}
$$

where we used equation (3.2) to derive the last equality. Integrating with respect to t and taking the conditional expectation at time t, we obtain

$$
\begin{aligned}
& \mathbf{E}_t^Q \left(e^{-r\,T}\, W(S_T,\, T+0) \right) \;-\; e^{-r\,t}\, W(S,\, t) \;\le \\
& - \mathbf{E}^Q \left\{ \sum_{t\,<\,T_i\,\le\,T} \lambda_i\, e^{-r\,T_i}\, G_i(S_{T_i}) \;-\; \int_t^T \eta\left(\sigma_s^2 \right)\, ds \right\}
\end{aligned}
$$

or, since , $W(S_T,\, T+0) \;=\; 0$,

$$
\mathbf{E}^Q \left\{ \sum_{t\,<\,T_i\,\le\,T} \lambda_i\, e^{-r\,(T_i\,-\,t)}\, G_i(S_{T_i}) \;-\; e^{r\,t} \int_t^T \eta\left(\sigma_s^2 \right)\, ds \right\} \;\le\; W(S,\, t)\ . \tag{A.5}
$$

This shows that the function $W(S,t)$ is an upper bound on the possible values taken by the left-hand side of (28) as Q ranges over the family of probabilities $\mathcal{P}$. The calculation also shows that the inequality becomes equality when the volatility of the Itô process is chosen to be precisely

$$
\sigma_t \;=\; \sqrt{\Phi'\left(\frac{e^{-r\,t}}{2}\, S_t^2\, W_{SS}\,(S_t,\, t) \right)} \tag{A.6}
$$

because the inequalities in (A.3) and (A.4) are saturated for this particular σ_t. Consequently, W is the value function for this control problem and (A.6) characterizes the measure where the supremum is attained.

A.2 Proof of Proposition 3.

We set

$$W_i \;=\; \frac{\partial W}{\partial \lambda_i} \quad \text{and} \quad W_{ij} \;=\; \frac{\partial^2 W}{\partial \lambda_i \, \partial \lambda_j}\;.$$

Differentiating equation (3.2) with respect to the variables λ_i, we obtain

$$\begin{aligned} & W_{i\,t} \;+\; \frac{1}{2}\,\Phi'\left(\frac{e^{-rt}}{2} S^2\, W_{SS}\right)\, S^2\, W_{i\,SS} \\ & +\; \mu\, S\, W_{i\,S} \;-\; r\, W_i \;=\; -\;\delta(t-T_i)\, G_i\;, \quad 1 \;\le\; i \;\le\; M\;, \end{aligned} \tag{A.7}$$

with $W_i(S,\,T+0) \;=\; 0$, and

$$\begin{aligned} & W_{ij\,t} \;+\; \frac{1}{2}\,\Phi'\left(\frac{e^{-rt}}{2} S^2\, W_{SS}\right)\, S^2\, W_{ij\,SS} \;+\; \frac{e^{-rt}}{4}\, S^4\, \Phi''\left(\frac{e^{-rt}}{2} S^2\, W_{SS}\right)\, W_{i\,SS}\, W_{j\,SS} \\ & +\; \mu\, S\, W_{ij\,S} \;-\; r\, W_{ij} \;=\; 0 \quad , \qquad 1 \;\le\; i,j \;\le\; M\;, \end{aligned} \tag{A.8}$$

with $W_{ij}(S,\,T+0) \;=\; 0$.[25]

Equation (A.7) describes the evolution of the gradient of W with respect to λ. It is a Black-Scholes-type equation in which the volatility parameter, depends on S and t.

The second equation has a "source term"

$$\frac{e^{-r\,t}}{4}\, S^4\, \Phi''\left(\frac{e^{-r\,t}}{2} S^2\, W_{SS}\right)\, W_{i\,SS}\, W_{j\,SS}$$

[25] The smoothness of the function $\Phi(X)$ justifies this formal differentiation procedure.

and no explicit dependence on the G_i's. To show that W is convex, it is sufficient to verify that, for all $\theta \in \mathbf{R}^M$, we have

$$H \;=\; \sum_{i=0,\,j=0}^{M} \theta_i\,\theta_j\,W_{ij} \;\geq\; 0\;. \tag{A.9}$$

But it follows from (A.8) that H satisfies

$$H_t \;+\; \frac{1}{2}\,\Phi'\left(\frac{e^{-r\,t}}{2} S^2\,W_{SS}\right)\,S^2\,H_{SS} + \frac{e^{-r\,t}}{4}\,S^4\,\Phi''\left(\frac{1}{2}S^2\,W_{SS}\right)\cdot\left(\sum_1^M \theta_i\,W_{i\,SS}\right)^2$$

$$+\;\mu\,S\,H_S \;-\; r\,H \;=\; 0 \tag{A.10}$$

with final condition $H(S,\,T) \;=\; 0$. Due to the convexity of Φ, we have

$$\frac{e^{-r\,t}}{4}\,S^4\,\Phi''\left(\frac{e^{-r\,t}}{2}S^2\,W_{SS}\right)\cdot\left(\sum_1^M \theta_i\,W_{i\,SS}\right)^2 \;\geq\; 0\;. \tag{A.11}$$

Hence, by the Maximum Principle applied to equation (A.10), we conclude that $H(S,\,t) \geq 0$ for all S and all $t < T$.

Finally, we show that $H(S,\,t) > 0$ to establish strict convexity. For this, it it sufficient to show that the left-hand side of (A.11) is positive on a subset of the $(S,\,t)$ plane of positive Lebesgue measure.[26] Notice that $\Phi''(X) > 0$ in a neighborhood of $X = 0$, which implies

$$\Phi''\left(\frac{e^{-r\,t}}{2}S^2\,W_{SS}\right) \;>\; 0$$

in regions of the $(S,\,t)$ plane where $\left|S^2\,W_{SS}\right|$ is sufficiently small. Recalling that the curvature of the payoffs decays as S tends to zero or infinity, we conclude that the latter inequality is satisfied if S is sufficiently far away from all strikes.

In addition, we note that

$$\sum_1^M \theta_i\,W_{i\,SS}$$

[26] This implies the positivity of H because the fundamental solution of equation (A.10) is strictly positive. Here we use the fact that $\Phi'(X) \geq \sigma_{min} > 0$.

cannot vanish on *any* open subset of the $(S,\,t)$ plane. Indeed, by the Unique Continuation Principle, this would imply that it vanishes identically and thus that

$$\sum_{1}^{M} \theta_i\, W_i$$

is a linear function. This is impossible unless

$$\sum \theta_i\, G_i$$

is linear at each payoff date, which is ruled out by the assumption that there is a single option per strike and at least one nonzero strike. We conclude that strict inequality holds in (A.11) for $(S,\,t)$ in some open set. Thus, W is strictly convex in $(\lambda_1,\, ...\lambda_M)$.

Appendix B: Dataset for the USD/DEM example

Maturity	Type	Strike	Bid	Offer	Mid	IVOL
	Call	1.5421	0.0064	0.0076	0.0070	14.9
	Call	1.5310	0.0086	0.0100	0.0093	14.8
30 days	Call	1.4872	0.0230	0.0238	0.0234	14.0
	Put	1.4479	0.0085	0.0098	0.0092	14.2
	Put	1.4371	0.0063	0.0074	0.0069	14.4
	Call	1.5621	0.0086	0.0102	0.0094	14.4
	Call	1.5469	0.0116	0.0135	0.0126	14.5
60 days	Call	1.4866	0.0313	0.0325	0.0319	13.8
	Put	1.4312	0.0118	0.0137	0.0128	14.0
	Put	1.4178	0.0087	0.0113	0.0100	14.2
	Call	1.5764	0.0101	0.0122	0.0112	14.1
	Call	1.5580	0.0137	0.0160	0.0149	14.1
90 days	Call	1.4856	0.0370	0.0385	0.0378	13.5
	Put	1.4197	0.0141	0.0164	0.0153	13.6
	Put	1.4038	0.0104	0.0124	0.0114	13.6
	Call	1.6025	0.0129	0.0152	0.0141	13.1
	Call	1.5779	0.0175	0.0207	0.0191	13.1
180 days	Call	1.4823	0.0494	0.0515	0.0505	13.1
	Put	1.3902	0.0200	0.0232	0.0216	13.7
	Put	1.3682	0.0147	0.0176	0.0162	13.7
	Call	1.6297	0.0156	0.0190	0.0173	13.3
	Call	1.5988	0.0211	0.0250	0.0226	13.2
270 days	Call	1.4793	0.0586	0.0609	0.0598	13.0
	Put	1.3710	0.0234	0.0273	0.0254	13.2
	Put	1.3455	0.0173	0.0206	0.0190	13.2

TABLE 1. Dataset used for the calibration example of Figs. 1, 2 and 3. Contemporaneous USD/DEM option prices (based on bid-ask volatilites and risk-reversals) provided to us by a marketmaker on August 23, 1995. The options correspond to 20-delta and 25-delta USD/DEM puts and calls and 50-delta calls. Implied volatility corresponding to mid-market prices for each option are displayed in the last column. The other market parameters are : spot FX = 1.4885/4890; DEM deposit rate= 4.27%; USD deposit rate= 5.91% .