

Computation of Deterministic Volatility Surfaces

Nicolas Jackson* Endre Süli
Oxford University Computing Laboratory,
Numerical Analysis Group, Wolfson building,
Parks Road, Oxford OX1 3QD

Sam Howison
Oxford Centre for Industrial and Applied Mathematics,
Mathematical Institute, 24-29 St Giles', Oxford OX1 3LB

26 October 1998

*I would like to acknowledge the financial support of the EPSRC and the investment bank BZW, and to thank Simon Acomb and Andrew Gray for helpful comments.

Abstract

The ‘volatility smile’ is one of the well-known biases of Black-Scholes models for pricing options. In this paper, we introduce a robust method of reducing this bias by pricing subject to a deterministic functional volatility $\sigma = \sigma(S, t)$. This instantaneous volatility is chosen as a spline whose weights are determined by a regularised numerical strategy that approximately minimises the difference between Black-Scholes vanilla prices and known market vanilla prices over a range of strikes and maturities; these Black-Scholes prices are calculated by solving the relevant partial differential equation numerically using finite element methods. The instantaneous volatility generated from vanilla options can be used to price exotic options where the skew and term-structure of volatility are important, and we illustrate the application to barrier options.

Contents

1	Introduction	4
2	Space-time spline representation of $\sigma(S, t)$	7
3	Regularised optimisation strategy	9
3.1	The cost-functional ϕ	9
3.2	Method of functional minimisation	12
4	Example volatility function generation	12
4.1	Implementation	12
4.2	Integrity testing	16
5	Example volatility function pricing: up-and-out call	18
6	Conclusions	19
	References	21
	Appendix	
A	Black-Scholes straddle prices	22
A.1	Differential problems	22
A.2	Discretisation method	23

1 Introduction

The standard Black-Scholes model (see for example [3], [10], [19]) for pricing options assumes that the underlying asset price S follows the stochastic differential equation

$$dS/S = (\mu - d)dt + \sigma dX \quad (1.1)$$

where μ is the expected growth rate, d is a constant continuous dividend yield, σ is the volatility and X is a standard Brownian motion. The model usually assumes that μ and σ are either constant or prescribed functions of time, although it is valid if σ is a deterministic function of S also. Following a no-arbitrage argument it is then shown that the option price V must satisfy the Black-Scholes equation

$$V_t + \frac{1}{2}\sigma^2 S^2 V_{SS} + (r - d)SV_S - rV = 0$$

where r is the continuously compounded risk-free interest rate. We may then price particular contracts by subjecting this equation to the relevant payoff (final-condition) and boundary conditions and solving the resulting differential problem. The hedging parameters (delta, gamma, etc.) can be found subsequently by differentiation, whether numerical or analytical, although special consideration should be made for rho and vega if either of r and σ are functions of S and/or t .

The choice of volatility σ (which measures the standard deviation of the rate of change of S) is crucial to the price. We may choose σ to be a constant estimated via a statistical analysis of historical S or, as is common market practice, we may choose σ as the *implied volatility* – the constant that satisfies the Black-Scholes equation if V is taken as a recent known market price of the option.

However, in practice, it is almost invariably found that the implied volatility is different from option to option, varying with both strike E and maturity T . This variation is known as the volatility smile, its spatial component being referred to as the *skew* of volatility, its temporal component as the *term structure*. It follows that no one constant σ can give prices consistent with market data.

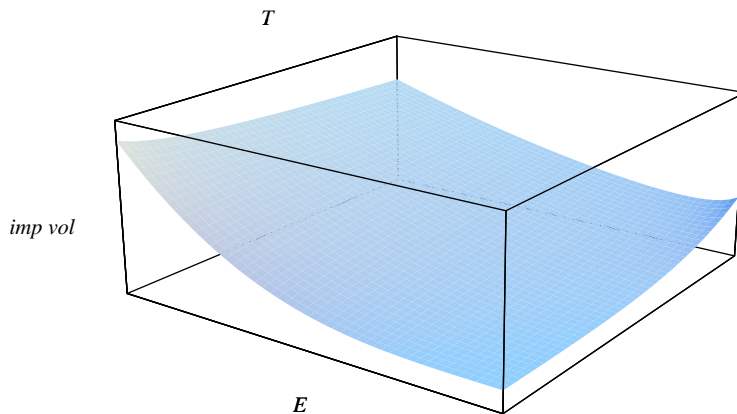


Figure 1: Schematic of a volatility smile.

While the use of a constant implied volatility may be adequate for liquid vanilla options it may be inappropriate for more exotic options where the effect of σ and the

bias of the volatility smile is more pronounced. Many remedies have been proposed for this situation. We focus on one of the simplest, that one should use a ‘deterministic’ functional volatility $\sigma(S, t)$ for pricing. This approach therefore stands in contrast to, say, stochastic volatility models (see, for example, [9], [8]) or jump diffusion models (see [15]) in which a second source of randomness is incorporated to explain better the bias in observed prices.

In this paper we present a method for finding a suitable representation of $\sigma(S, t)$ by minimising an appropriate measure of the difference between Black-Scholes vanilla prices and known market vanilla prices over a range of strikes and maturities. Assuming that the asset price does in fact evolve according to (1.1), we will find a form for $\sigma(S, t)$ with the following desirable properties:

- It will almost remove the bias of the volatility smile, in that calculated vanilla prices will be close to recent known market vanilla prices.
- A probability density function (pdf) for future values of S implied by market vanilla prices (see [13] for a method of construction), over a range of strike prices and maturities, will approximately coincide with the pdf for future values of S subject to the functional volatility. We shall not demonstrate this property but it is implied by our method for constructing $\sigma(S, t)$.
- Since the functional form will be chosen by consideration of recent known market vanilla prices, it can be used to price exotic options consistently with the hedging instruments.

It should be noted though, that in using vanilla prices to determine $\sigma(S, t)$, we abandon any attempt to predict these prices (although we can of course predict their greeks). Thus, for example, if the asset price evolves according to some process other than (1.1) with a deterministic volatility, we do not attempt to find any mispricings among vanilla products, but rather take them as given. To this extent the deterministic volatility model is less ambitious than some others, such as those mentioned above. It should be noted, however, that the necessity for parameter estimation makes the other models less practical.

In [6], Dupire shows that if vanilla option implied volatility were available for all conceivable strikes and maturities then $\sigma(S, t)$ is uniquely determined. However the smile is not defined everywhere in practice (because only a finite number of strikes and maturities are traded), so $\sigma(S, t)$ is not uniquely determined. Moreover, the numerical ‘inverse problem’ of choosing a suitable form for $\sigma(S, t)$ is often unstable even if we extend the smile to a suitably smooth continuous surface through interpolation. Therefore the problem of finding $\sigma(S, t)$ is naturally ill-posed.

Dupire [6], Derman & Kani [5] and Rubinstein [16] have each presented implied binomial/trinomial tree algorithms for pricing options that use the degrees of freedom of their lattices so as to fit the smile (or rather an interpolated extension of the discrete array of implied volatilities) to the price. The construction of a functional form for σ is thus implicit in these algorithms. They have the advantage of relatively fast execution but they have several disadvantages:

- Algorithm [16] assumes that volatility only has skew structure, i.e. that $\sigma = \sigma(S)$. However, the term-structure of volatility may be important for many path-dependent options such as Americans or barriers and it is inconsistent to

use different $\sigma(S)$ for different maturities. Jackwerth [12] has addressed this difficulty but at the cost of a large-scale optimisation problem.

- Algorithms [6] and [5] do not effectively counter the natural ill-posedness of the problem of ‘fitting the smile’ so they are not robust. Unusual behaviour in the smile or too high interest rates can cause the algorithms to fail or to allow negative probabilities (cf. Barle & Cakici [2] who attempt to reduce these difficulties).
- Lattice methods only price for a single spot price and not for a range. More seriously, lattice methods only determine the instantaneous volatility at lattice points, which cover only a (triangular) proportion of the whole (rectangular) (S, t) domain. In pricing exotics, in particular, values of $\sigma(S, t)$ outside this triangular domain may be important.
- More generally, it is difficult to apply such algorithms to barrier option or multi-dimensional pricing problems.

More recently, Lagnado & Osher [14] have presented a regularised minimisation method, using finite difference methods, to fit the smile to option prices. Regularisation is here a self-stabilising procedure in the minimisation process (see for example [18]) which restricts σ to the smoothest functions that minimise the difference between Black-Scholes prices and recent known market prices. This attempts to force well-posedness of the ‘inverse problem’. Lagnado & Osher’s method appears to be effective and robust; however, it has some disadvantages:

- The method is particularly computationally expensive – the calculation of variational derivatives in their gradient descent minimisation requires extensive solution of partial differential problems.
- The method only generates a discrete representation of the instantaneous volatility function described by a relatively small array of nodes in space-time. This is disadvantageous because a knowledge of σ away from these nodes is often required in pricing exotics. There is no unique way to interpolate this array, and different (sensible) interpolants may yield significantly different option prices.
- The method gives equal importance to all known market prices (or equivalently all implied volatilities) while it may be better to give greater weight to options that are more heavily traded, shorter dated or nearer to the money.

We also mention a recent paper of Avellaneda et al. [1] who construct a representation for $\sigma(S, t)$ via a relative-entropy minimisation method. The instantaneous volatility that their method constructs has particularly sharp peaks and troughs near strike/expiration dates. There are two possible drawbacks to such an outcome. One is that it may be unrealistic – there may be no obvious a priori reason why volatility should change abruptly just because a particular option contract expires. The other is numerical. While accurate numerical solution of a Black-Scholes problem (using this irregular volatility) may replicate a given set of market prices, coarse numerical solutions are liable not to see the abrupt changes and thus not replicate the prices to sufficient accuracy. In practice, due to critical financial time-constraints, it is often necessary to discretise the space-time domain coarsely in a lattice or finite difference/element pricing method, and therefore it is advantageous for the instantaneous

volatility to be as smooth as possible in nature so as to retain accuracy. Similar remarks apply to Monte Carlo pricing schemes.

In this paper we present a robust algorithm for generating a near-optimal choice for $\sigma(S, t)$ via a regularised minimisation strategy. We choose the instantaneous volatility function to be a space-time spline; this choice is motivated by careful experimentation. Our representation of $\sigma(S, t)$ is therefore guaranteed to be a smooth function (the simplest – and smoothest – functional volatility is a single constant implied volatility!). We consider it to be reasonable to restrict ourselves to smooth volatility functions as we have no reason to believe that today’s market prices can forecast abrupt changes in future volatility with any reliability whatsoever. We choose the constants that uniquely determine this spline via a regularised and weighted minimisation of the difference between Black-Scholes prices and known market prices over a range of strikes and maturities. We solve the Black-Scholes differential problem in the minimisation process using a finite element method.

We independently analyse the integrity of our method using the adaptive finite element method presented in [11]. This adaptive finite element method solves European Black-Scholes problems to guaranteed accuracy, and thus it may be used to verify that a generated volatility function does indeed yield vanilla option prices that match the given market vanilla option prices to within a prescribed error tolerance. Such integrity testing is important because it is all too easy to generate a representation for instantaneous σ via a discretised tree or finite difference/element method that appears to price a set of vanilla options correctly, only to find that a much finer discretisation yields a significantly different set of prices.

It is particularly advantageous to be confident that the generated instantaneous σ is not tied to discretisation errors of a particular generating method (whether that be an implied-tree or finite difference/element optimisation method) and is continuously well-defined in space-time since it is then not method-dependent, and thus can be used consistently by many pricing methods (for example Monte Carlo simulation).

In Section 2 we describe the space-time spline for σ . In Section 3 we describe the functional and minimisation method that are used to configure the constants that uniquely define the spline. An example of instantaneous volatility generation is given in Section 4: given an example set of recent known market straddle option prices on the FTSE-100, we construct σ and then prove that it does effectively remove the bias of the volatility smile. An example pricing of an exotic option using the generated σ is given in Section 5. Finally, in Section 6, we conclude the work presented herein.

2 Space-time spline representation of $\sigma(S, t)$

It is advantageous to represent $\sigma(S, t)$ by a space-time spline for the following reasons:

- It is simple and convenient to have a structure that is uniquely determined by a finite number of constant weights, this number being user-configurable.
- We then know that σ has the smoothness properties that we prescribe; we require σ to have at least one continuous spatial derivative and to be at least continuous in time.
- Without a priori knowledge of a suitable form for σ , it allows local functional variation and thus is reasonably general. Moreover, by increasing the number

of splines used, complicated functional forms can be represented.

- We find that simple non-spline representations (for example $\sigma(S) = aS^{-b}$ for constants $a > 0$ and $b \in [0, 1]$, the CEV model of Cox & Ross [4]) are usually not general enough to perform well. More complicated non-spline representations may perform well in certain circumstances but they lack the flexibility of a spline.

For each t , we describe the ‘spatial’ S -variation of $\sigma(S, t)$ by a P -dimensional *natural cubic spline* (see, for example, Schwarz [17]) for $S \in [S_{min}, S_{max}]$; we choose $S_{min} > 0$ and $S_{max} < \infty$ to approximate 0 and ∞ (see Appendix A.1). Then, for each t , σ is uniquely determined by specifying its value at a set of $P + 1$ nodes $\cup_p \{S_p\}$, where $S_{min} = S_0 < S_1 < \dots < S_p < \dots < S_P = S_{max}$. For simplicity the set of asset price nodes S_p are not time-varying and are chosen upfront. They need not coincide with any particular spot or strike prices. However, it is sensible to choose a higher proportion of the S_p in the financial region of interest, near to at-the-money.

The natural cubic spline is the smoothest of all piecewise cubic interpolants in the sense that the energy functional $\frac{1}{2} \int_{S_{min}}^{S_{max}} \sigma_{SS}^2 dS$ is minimised. For each t , having specified the value of σ at each S_p , spatial derivatives of σ are chosen uniquely and automatically by the construction method [17] to ensure that σ is C^2 , i.e. that σ has two continuous S -derivatives. In Figure 2 we graph a typical σ against S for some time t ; the alternate shades denote different cubics. The vertical lines delineate the region near to at-the-money, in which we have represented σ by six cubics, while deeply away from at-the-money there are just four.

We now turn to the time dependence. We choose a selection of $Q + 1$ temporal nodes $0 = t_0 < t_1 < \dots < t_q < \dots < t_Q = T_{max}$, and at each spatial node S_p we force $\sigma(S_p, t)$ to be piecewise linear in time. These temporal nodes are chosen upfront and need not coincide with any particular expiry times.

Thus the approximation of $\sigma(S, t)$ is specified by the matrix of weights

$$\Sigma = \begin{pmatrix} \Sigma_{00} & \Sigma_{10} & \dots & \Sigma_{P0} \\ \Sigma_{01} & \Sigma_{11} & & \Sigma_{P1} \\ \vdots & & \ddots & \vdots \\ \Sigma_{0Q} & & & \Sigma_{PQ} \end{pmatrix}^T$$

where $\Sigma_{pq} = \sigma(S_p, t_q)$. This determines the representation uniquely.

In Figure 3 we extend the schematic of Figure 2 (with $P = 10$) to sketch a sample term-structure (with $Q = 3$). We note that σ is linear in t only at the nodes S_p . Therefore away from these nodes strictly speaking it does not have a piecewise linear spline representation in time.

In due course we shall present a method for selecting the members of Σ by a minimisation method. However, we propose that the members of Σ that fall on or near to $S = S_{min}$ and $S = S_{max}$ not be part of this minimisation directly. We will instead ‘link’ these asymptotic values to those being optimised. More precisely, we will only optimise over a submatrix Σ^* of Σ that contains all of the entries of Σ except some of its first and last rows; the entries of the remaining first row(s) of Σ will be set to be always equal to the entries of the first row of Σ^* ; the entries of the remaining last row(s) of Σ will be set to be always equal to the entries of the last row of Σ^* . So specifying Σ^* determines the complement $\Sigma \setminus \Sigma^*$.

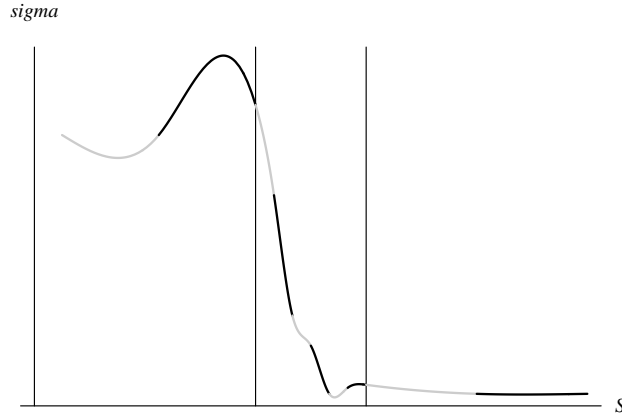


Figure 2: Schematic for spatial variation of σ .

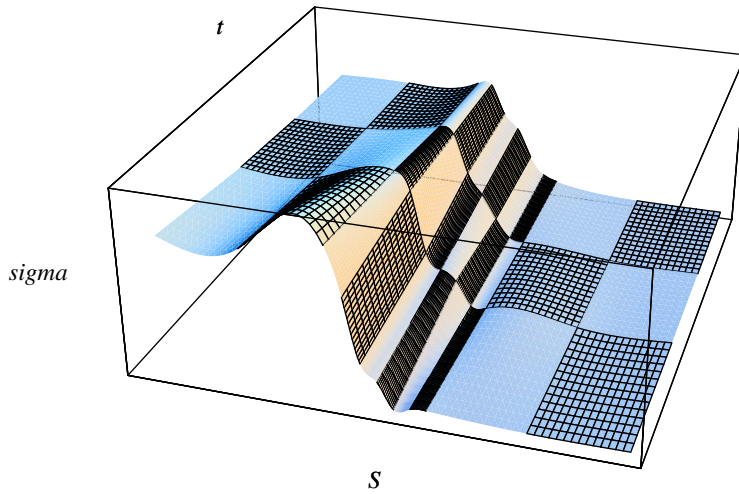


Figure 3: Schematic for space-time variation of σ .

We adopt this procedure because the values of σ far from at-the-money do not significantly effect option prices, and because it considerably reduces the computational complexity of the minimisation, which leads to greater efficiency.

3 Regularised optimisation strategy

3.1 The cost-functional ϕ

It is financially appropriate to minimise some measure of the difference between Black-Scholes vanilla option prices V (calculated assuming a functional volatility) and known market vanilla option prices \bar{V} (the arithmetic average of recent known bid and offer prices) over a range of strike prices E_i , for $i = 1, \dots, I$, and maturities T_j , for $j = 1, \dots, J$, relative to a given spot price $S = S(0)$. Naturally we might select a different spot price (and time) for each pair (E_i, T_j) , corresponding to the spot price (and time) at which the trade was made, but here it is convenient to work with just one. The range of strike prices and maturities need not be described by a rectangular

array but here they are for convenience. We specifically select these vanilla options to be straddles, that is portfolios long one call and long one put of the same strike and maturity. Straddles have been chosen to represent vanilla options because of their ability to preserve put-call parity more accurately in the numerical procedure that follows.

We might attempt to find the $\sigma(S, t)$ that minimises the weighted discrete L^2 norm

$$F := \left(\sum_{i,j} w_{ij} |\bar{V}_{ij} - V_{ij}|^2 \right)^{1/2} \quad (3.1)$$

with a weighting function w that reflects the relative importance/significance of the particular market straddle prices; we suppose $w_{ij} > 0$ and $\sum_{i,j} w_{ij} = 1$. The Black-Scholes prices V_{ij} , subject to a particular volatility $\sigma(S, t)$, are obtained by solving the relevant partial differential problem numerically, and we present our finite element method for this purpose in Appendix A.

However, because I and J are finite, the problem of minimising F is not well-posed as typically it has many solutions. It is straightforward to demonstrate this computationally and we do so in Section 4.1. By well-posedness we mean that there exists a unique σ , that has a continuous data-dependence and is contained in an admissible space of functions, such that F is minimal.

Therefore we *regularise* our strategy by instead choosing to minimise

$$\phi := F^2 + G^2, \quad (3.2)$$

where we define

$$G^2 := \sum_{p,q} \frac{c_1 \sigma_S^2(S_p, t_q)}{(P+1)(Q+1)} + \sum_{\substack{p,q \\ q \neq Q}} \frac{c_2 \sigma_t^2(S_p, t_q^+)}{(P+1)Q}, \quad (3.3)$$

with the subscripts to σ denoting partial derivatives, for positive constants c_1 and c_2 . Clearly G^2 is a measure of the size of the average gradient.

Also, we constrain σ to satisfy $\sigma_{min} \leq \sigma(S, t) \leq \sigma_{max}$ over the space-time domain, where σ_{min} and σ_{max} are chosen constants such that $0 < \sigma_{min} < \sigma_{max} < \infty$. The financial reason for constraining σ to bounded positive values is clear but it is also important numerically that σ_{min} is as large as possible. This is because when $\sigma \ll 1$ (for example, if $\sigma = 0.00001$ at some point in space-time) the Black-Scholes equation becomes increasingly convection-dominated and hyperbolic in nature which makes accurate solution of the Black-Scholes problem difficult.

The new term G^2 has two effects as far as we are concerned. The first is to force uniqueness of σ if c_1 and c_2 are large enough. This is found to be the case experimentally although we do not attempt to prove it; we discuss the details in Section 4.1. The second effect is to make σ increasingly smooth as c_1 and c_2 are chosen to be larger and larger; in the limit $c_1, c_2 \rightarrow \infty$, the minimising σ is a constant, the ‘best-fit implied volatility’. Consequently, if c_1 and c_2 are very large, calculation with the resulting σ may give prices that differ noticeably from market prices.

These considerations of uniqueness and smoothness again force us to face the question: what properties should σ have? Why should we ask for uniqueness when

we know that it cannot be expected if we only have finite number of market prices? Of several different choices for σ that all reproduce market prices closely enough, why should we choose the smoothest? One practical reason for specifying uniqueness is that the model may need to be recalibrated frequently as market data changes, so that if we were to use σ to price exotic derivatives, or to hedge products, then a switch to a significantly different σ may predict unreasonably large changes in prices or hedge ratios. The choice of a smoother σ has two potential advantages. The first is numerical: the smoother σ is, the faster and more accurate certain numerical schemes are. The second is more ad hoc: we may take the view that it is not appropriate to force calculated option prices V_{ij} to be too close to market prices \bar{V}_{ij} on the grounds that some of the latter may be out of line, or misquoted. Taking c_1 and c_2 to be very large forces σ to be nearly constant, and consequently allows larger differences between calculated and market prices, which may be arbitrage opportunities. (An alternative approach to ‘outliers’ is to try to identify them either before or after the minimisation, reject them, and calculate or recalculate σ using a reduced set of market prices.) It is an interesting empirical question whether trading on this basis yields noticeably different results than trading a volatility surface that matches all market prices very closely.

Some care is thus needed in the choice of c_1 and c_2 . If we aim to select them so as to ensure σ is unique, it is not immediately apparent how large they should be. However, their rough order of magnitude can be found as follows. It is reasonable to aim for roughly similar magnitudes of F^2 and G^2 at the minimum of ϕ , subject to the constraints; if G^2 were to dominate then the deterministic volatility surface would be artificially flat, while if F^2 were to dominate we might encounter the problems of non-uniqueness mentioned above. (It may be prudent to check a posteriori that the gradients of the deterministic volatility surface are of roughly similar magnitude to those of the implied volatility surface.) If $F = F_{min}$ at this minimum, and if δS and δt are typical scales for variations in S and t , then

$$F_{min}^2 \sim G_{min}^2 \sim c_1 \left(\frac{\delta\sigma_1}{\delta S}\right)^2 + c_2 \left(\frac{\delta\sigma_2}{\delta t}\right)^2,$$

where $\delta\sigma_1$ is a typical variation in σ that results from $S \mapsto S \pm \delta S$ and $\delta\sigma_2$ is a typical variation in σ that results from $t \mapsto t \pm \delta t$. Thus, on balancing these terms, we might assume

$$c_1 = \bar{c} \left(\frac{\delta S F_{min}}{\delta\sigma_1}\right)^2, \quad c_2 = \bar{c} \left(\frac{\delta t F_{min}}{\delta\sigma_2}\right)^2,$$

for a single regularising constant \bar{c} ; the choice of \bar{c} is then found experimentally, though it is expected to be approximately $\mathcal{O}(1)$. Concerning the value of F_{min} that we should aim for, it should certainly not be smaller than the errors associated with the numerical method, and it may be much larger if we aim to identify other discrepancies such as mispricings. The δ -parameters are just rough order-of-magnitude estimates: if $\bar{\sigma}$ denotes a typical at-the-money implied volatility for a typical maturity, and if \bar{T} denotes a typical maturity, we choose $\delta S = S(0)\bar{\sigma}\sqrt{\bar{T}}$ (the standard deviation of changes in S over time \bar{T} under constant volatility) and $\delta t = \bar{T}$; we select $\delta\sigma_1$ and $\delta\sigma_2$ using the typical gradients of the implied volatility matrix. It is not important to select these values precisely as in any case \bar{c} must still be chosen empirically.

3.2 Method of functional minimisation

We employ the quasi-Newton algorithm of Gill & Murray [7] to minimise ϕ subject to the linear constraints $\sigma_{min} \leq \sigma(S, t) \leq \sigma_{max}$; we only apply the constraints at each of the nodes of the splined σ for convenience (strictly speaking this will not ensure the constraints are satisfied over the whole space-time domain but, provided σ_{min} is not too small, this is not problematic). In fact we only optimise over a submatrix Σ^* of Σ , as described in Section 2, since it is appropriate to ‘link’ those asymptotic members of Σ that lie on or near to $S = S_{min}$ and $S = S_{max}$ to members of Σ^* .

From a starting approximation $\Sigma^{*(0)}$ to the solution matrix Σ^* , such that

$$\phi(\Sigma^*) = \min \phi$$

subject to the constraints, on the basis of estimates of the gradient and curvature of ϕ , the algorithm generates a sequence of feasible matrices that are intended to converge to Σ^* . The algorithm may be terminated either when the number of quasi-Newton iterations has reached an unacceptable maximum or when changes in the approximations to Σ^* are small, for example when

$$\left(\frac{1}{(P^* + 1)(Q + 1)} \sum_{p=0}^{P^*} \sum_{q=0}^Q |\Sigma_{pq}^{*(\nu)} - \Sigma_{pq}^{*(\nu-1)}|^2 \right)^{1/2} < TOL, \quad (3.4)$$

for a user-specified constant TOL , where $\Sigma^{*(\nu)}$ is the ν th approximation to Σ^* . Alternatively, we could adopt the stopping criterion that

$$\sum_{i,j} w_{ij} |\mathcal{E}_{ij}| < TOL', \quad (3.5)$$

for a user-specified constant TOL' , where we define the basis-point error

$$\mathcal{E}_{ij} := 10^4 (\bar{V}_{ij} - V_{ij}) / S(0).$$

That is, we stop when we have reproduced market prices to a desired accuracy. There are some disadvantages to this approach, however. One is that it may not be possible to satisfy (3.5) for a given TOL' , especially if c_1 and c_2 are large. Furthermore, it is not consistent with the gradient regularisation, but rather with simply minimising F^2 (not $F^2 + G^2$). Thus, while TOL' may be large enough to allow the iteration to stop, the approximated Σ^* may not be close to the true minimiser; in particular the associated volatility σ may be less smooth.

4 Example volatility function generation

4.1 Implementation

Suppose we wish to generate instantaneous σ for the FTSE-100 index so as to price exotic options today. Then we might select a set of yesterday’s known market straddle prices over a range of five strike prices and two maturities, i.e. ten market prices in all with respect to the same spot price of $S(0) = 5000$. For this example we have chosen the sets

$$(E)_i = (4800 \quad 4900 \quad 5000 \quad 5300 \quad 6000),$$

$$(T)_j = (0.5 \quad 1.0)^T$$

and

$$(\bar{V})_{ij} = \begin{pmatrix} 429.40 & 384.95 & 353.98 & 388.16 & 934.14 \\ 637.36 & 601.38 & 573.85 & 560.29 & 892.26 \end{pmatrix}^T.$$

We measure time in years. We are required to specify the continuously compounded risk-free rate for each maturity (obtained from the discount factors of zero-coupon bonds) and to model the effects of dividends for each maturity (the dividends themselves being known or forecasted). Here, we shall assume a constant r and a constant continuous dividend yield d for each T_j for convenience; more generally we could assume r to be a known function of time and assume a discrete dividend model. Accordingly, the continuously compounded risk-free rate and constant continuous dividend yield vectors are taken to be $(r)_j = (0.04974 \quad 0.05354)^T$ and $(d)_j = (0.032 \quad 0.027)^T$.

There are $I = 5$ strikes and $J = 2$ maturities while as yet an unspecified number of weights $(P + 1) \times (Q + 1)$ are used to describe the splined σ . We shall take $P = 9$ and $Q = 2$, with $S_{min} = 500$ and $S_{max} = 50000$, and with the spline nodes

$$(S)_p = (S_{min} \quad 3000 \quad 4500 \quad 4750 \quad 5000 \quad 5300 \quad 6000 \quad 10000 \quad 25000 \quad S_{max})$$

and

$$(t)_q = (0.0 \quad 0.5 \quad 1.0)^T.$$

We prescribe two columns of asymptotic values for Σ^T on or near to $S = S_{min}$, and three columns of asymptotes on or near to $S = S_{max}$; that is we have

$$\Sigma = \begin{pmatrix} \Sigma_{00}^* & \Sigma_{00}^* & & \Sigma_{04}^* & \Sigma_{04}^* & \Sigma_{04}^* \\ \Sigma_{10}^* & \Sigma_{10}^* & \Sigma^{*T} & \Sigma_{14}^* & \Sigma_{14}^* & \Sigma_{14}^* \\ \Sigma_{20}^* & \Sigma_{20}^* & & \Sigma_{24}^* & \Sigma_{24}^* & \Sigma_{24}^* \end{pmatrix}^T.$$

To begin with we use (3.4) as the stopping criterion for the minimiser with the highly restrictive tolerance $TOL = 5 \times 10^{-(1+\alpha)} / \sqrt{(P^* + 1)(Q + 1)}$ so that ϕ , subject to the constraints, can be minimised very accurately by choosing α large enough. This is useful for analysis that follows since it implies that $|\Sigma_{pq}^{*(\nu)} - \Sigma_{pq}^{*(\nu-1)}| = 0$ to α -decimal accuracy for all p and q . We choose $\alpha = 6$. Space-time of the differential problems involved in the minimisation is discretised by $M = 60$ piecewise quadratic finite elements and $(N)_j = (30 \quad 40)^T$ time-steps (see Appendix A.2); we choose $N_1 \neq N_2$ here as it is more efficient to use a different number of time-steps for options of different maturities. We set $\sigma_{min} = 0.01$ and $\sigma_{max} = 1.0$. We choose the set of weights

$$(w)_{ij} = \frac{1}{54} \begin{pmatrix} 2 & 5 & 8 & 5 & 2 \\ 4 & 7 & 10 & 7 & 4 \end{pmatrix}^T.$$

Suppose we aim for $F_{min} = 1.5$. We choose $\bar{\sigma} = 0.1475$, $\bar{T} = 0.5$ and $\delta\sigma_1 = \delta\sigma_2 = 0.04$. Then $c_1 \approx 3.8(10^8)\bar{c}$ and $c_2 \approx 3.5(10^2)\bar{c}$. Then, with the starting approximation

$$\Sigma^{*(0)} = \begin{pmatrix} 0.15 & 0.15 & 0.15 & 0.15 & 0.15 \\ 0.15 & 0.15 & 0.15 & 0.15 & 0.15 \\ 0.15 & 0.15 & 0.15 & 0.15 & 0.15 \end{pmatrix}^T,$$

application of the algorithm to generate the near-optimal choice for σ yields the approximations

$$\Sigma^*(\bar{c} = 0.00) \approx \begin{pmatrix} 0.081494 & 0.070511 & 0.118491 & 0.086233 & 0.058517 \\ 0.234544 & 0.248359 & 0.133653 & 0.125910 & 0.072447 \\ 0.210362 & 0.242019 & 0.170321 & 0.164311 & 0.162403 \end{pmatrix}^T, \quad (4.1a)$$

$$\Sigma^*(\bar{c} = 0.01) \approx \begin{pmatrix} 0.153400 & 0.171052 & 0.109226 & 0.093051 & 0.078747 \\ 0.197047 & 0.215270 & 0.129533 & 0.125311 & 0.090951 \\ 0.234645 & 0.253174 & 0.174680 & 0.164172 & 0.154288 \end{pmatrix}^T, \quad (4.1b)$$

$$\Sigma^*(\bar{c} = 0.07) \approx \begin{pmatrix} 0.165628 & 0.169917 & 0.110937 & 0.082097 & 0.087034 \\ 0.204074 & 0.208146 & 0.129789 & 0.127549 & 0.098777 \\ 0.233750 & 0.233882 & 0.172702 & 0.161416 & 0.148681 \end{pmatrix}^T, \quad (4.1c)$$

$$\Sigma^*(\bar{c} = 0.10) \approx \begin{pmatrix} 0.172465 & 0.167995 & 0.111406 & 0.082620 & 0.088950 \\ 0.207248 & 0.205021 & 0.129416 & 0.126414 & 0.100498 \\ 0.232003 & 0.229633 & 0.172764 & 0.161788 & 0.147995 \end{pmatrix}^T. \quad (4.1d)$$

However, with the starting approximation

$$\Sigma^{*(0)} = \begin{pmatrix} 0.16 & 0.15 & 0.12 & 0.11 & 0.10 \\ 0.20 & 0.18 & 0.13 & 0.12 & 0.11 \\ 0.24 & 0.23 & 0.20 & 0.18 & 0.17 \end{pmatrix}^T, \quad (4.2)$$

application of the algorithm yields the approximations

$$\Sigma^*(\bar{c} = 0.00) \approx \begin{pmatrix} 0.010000 & 0.112017 & 0.120900 & 0.105606 & 0.179640 \\ 0.244730 & 0.230310 & 0.127910 & 0.113607 & 0.102161 \\ 0.190260 & 0.238494 & 0.173239 & 0.168179 & 0.154105 \end{pmatrix}^T, \quad (4.3a)$$

$$\Sigma^*(\bar{c} = 0.01) \approx \begin{pmatrix} 0.145859 & 0.181633 & 0.118454 & 0.081616 & 0.085109 \\ 0.195128 & 0.205182 & 0.124707 & 0.126128 & 0.097245 \\ 0.236547 & 0.253126 & 0.177269 & 0.164127 & 0.150755 \end{pmatrix}^T, \quad (4.3b)$$

$$\Sigma^*(\bar{c} = 0.07) \approx \begin{pmatrix} 0.165333 & 0.169889 & 0.111356 & 0.082090 & 0.086870 \\ 0.203874 & 0.207837 & 0.129517 & 0.127379 & 0.098628 \\ 0.233958 & 0.234162 & 0.172934 & 0.161600 & 0.148825 \end{pmatrix}^T, \quad (4.3c)$$

$$\Sigma^*(\bar{c} = 0.10) \approx \begin{pmatrix} 0.172465 & 0.167994 & 0.111406 & 0.082616 & 0.088952 \\ 0.207250 & 0.205020 & 0.129416 & 0.126415 & 0.100500 \\ 0.232001 & 0.229631 & 0.172763 & 0.161787 & 0.147994 \end{pmatrix}^T. \quad (4.3d)$$

In accordance with the asymptotic linking procedure, for each of these solution matrices Σ^* we have

$$\Sigma = \begin{pmatrix} \Sigma_{00}^* & \Sigma_{00}^* & & \Sigma_{04}^* & \Sigma_{04}^* & \Sigma_{04}^* \\ \Sigma_{10}^* & \Sigma_{10}^* & \Sigma^{*T} & \Sigma_{14}^* & \Sigma_{14}^* & \Sigma_{14}^* \\ \Sigma_{20}^* & \Sigma_{20}^* & & \Sigma_{24}^* & \Sigma_{24}^* & \Sigma_{24}^* \end{pmatrix}^T;$$

the same applies for the starting approximations $\Sigma^{*(0)}$.

If Σ^* were unique then it would be independent of its starting approximation. Therefore the matrices (4.1a) and (4.3a) provide a convincing demonstration of ‘non-uniqueness without regularisation’. Taken together, the sets of matrices (4.1) and

(4.3) indicate ‘uniqueness with sufficient regularisation’. Furthermore, since we have non-dimensionalised c_1 and c_2 , it is reasonable to suppose that $\bar{\tau} \approx 0.1$ is a good choice of regularising constant independently of the underlying.

The highly restrictive stopping criterion used above will make the minimiser very slow – it is also unnecessary other than for analysis. If we were to use the same criterion except with $\alpha = 1$ then, using as the starting approximation (4.2), we obtain

$$\Sigma^*(\bar{\tau} = 0.1) \approx \begin{pmatrix} 0.178160 & 0.150367 & 0.116918 & 0.101893 & 0.071722 \\ 0.206947 & 0.202562 & 0.124008 & 0.115095 & 0.093785 \\ 0.237264 & 0.231821 & 0.178387 & 0.169686 & 0.155311 \end{pmatrix}^T, \quad (4.4)$$

which is in reasonable agreement with (4.3d) and was much less expensive to compute.

If we had instead used (3.5) as the stopping criterion, with $TOL' = 2$ measuring an average number of basis-points error, then, using as the starting approximation (4.2), we obtain

$$\Sigma^*(\bar{\tau} = 0.1) \approx \begin{pmatrix} 0.211880 & 0.147707 & 0.117241 & 0.099936 & 0.052162 \\ 0.197152 & 0.192046 & 0.126953 & 0.113431 & 0.087933 \\ 0.232217 & 0.222051 & 0.186165 & 0.168521 & 0.164204 \end{pmatrix}^T, \quad (4.5)$$

which does not agree with (4.3d) as well as (4.4) does (consider the first columns of the transposed matrices), though it does give Black-Scholes prices V_{ij} that agree closely with market prices \bar{V}_{ij} .

For the remainder of this paper, we use the Σ determined by (4.4). This specifies the splined σ which is depicted in Figure 4. (The non-monotonic behaviour for small and large S is a result of the spline representation and our procedure for the asymptotic values $\Sigma \setminus \Sigma^*$. As stated previously, the consequences for pricing are insignificant.) While σ is defined for a large range of spot price, $500 \leq S \leq 50000$, the region nearest to at-the-money is of most financial interest and hence we depict this range specifically in Figure 5. Relative to the given set of market straddle prices, continuously compounded risk-free rates and constant continuous dividend yields, this functional form for the volatility may now be used to price many types of options on the FTSE-100, for example Americans, Asians, barriers, lookbacks etc..

Note that it is appropriate to retain at least three-decimal accuracy in Σ^* despite the possibility of low-accuracy market data. This is because the shape of the space-time spline, being entirely specified by the entries of Σ , is sensitive to each Σ_{pq}^* . Consider, if we rounded Σ^* to three-decimal accuracy then the average basis-point error, $\sum_{i,j} w_{ij} |\mathcal{E}_{ij}|$, would increase only slightly while rounding to two-decimal accuracy will more than double the error. If market data is deemed inaccurate then, rather than not retaining decimal accuracy, it would be appropriate to generate another Σ^* for a perturbed set of market data and then to account for the pricing differences implied by each of the Σ^* .

The minimiser took around 60 seconds to run on a SUN SPARC-station HS14 computer. Parallelisation of the minimiser would markedly reduce this time; a larger (smaller) set of market straddle prices will obviously imply more (less) run-time; more (fewer) splines will obviously imply more (less) run-time. The quality of the starting approximation $\Sigma^{*(0)}$ to Σ^* will greatly influence the run-time. If we generated Σ^* for

the FTSE-100 on a daily basis then we would expect a good starting approximation to be available from yesterday's calculation, the use of which will reduce run-time.

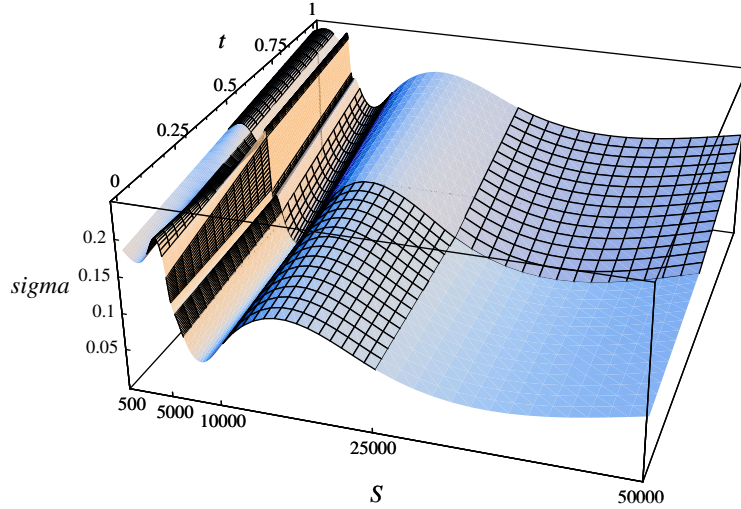


Figure 4: Today's instantaneous σ for the FTSE-100.

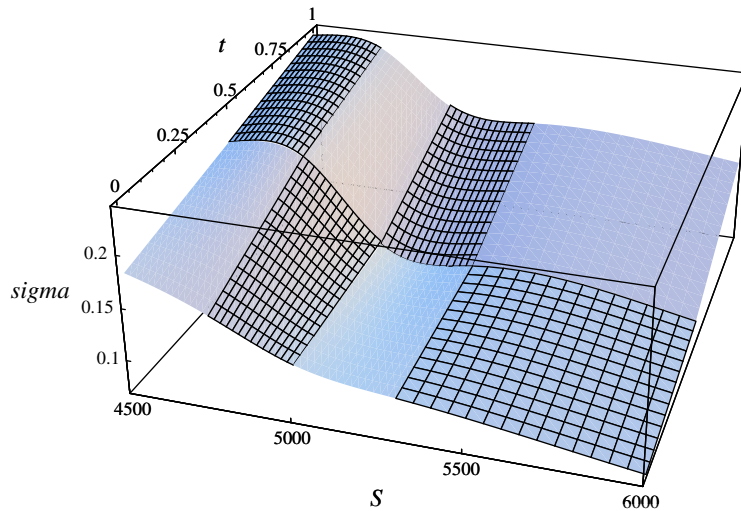


Figure 5: Today's instantaneous σ for the FTSE-100 (near to at-the-money).

4.2 Integrity testing

Recall that the Black-Scholes prices, V_{ij} , are approximated via numerical solution of the relevant partial differential problems, as described in Appendix A. Because we in fact approximate V_{ij} by $V_{ij} + e_{ij}$, with a numerical error e_{ij} , so we approximate F by

$$F + e^* = \left(\sum_{i,j} w_{ij} |\bar{V}_{ij} - V_{ij} - e_{ij}|^2 \right)^{1/2} .$$

Consequently, the smallness of $F + e^*$ (which is implied by the smallness of $\phi + e^{**}$, where e^{**} is the associated error in ϕ) need not imply the smallness of F ; conversely, $F + e^*$ may be much larger than F . Therefore it is important that $|e^*|$ should be small enough not to effect the minimisation significantly. Unless we adequately control the size of the error e_{ij} , our strategy for selecting instantaneous σ will not be effective as it will not ensure that the Black-Scholes prices, V_{ij} , are sufficiently close to the market prices, \bar{V}_{ij} .

The size of e_{ij} is directly related to the size of M and the N_j , which describe the coarseness of the discretisation of the relevant partial differential problems: the larger M and the N_j are, the smaller $|e_{ij}|$ will be, and vice-versa. Thus we are required to select M and the N_j carefully, for if they are too small then $|e_{ij}|$ will be too large while if they are too large then the minimiser will be inefficient and prohibitively slow! It is reasonable to expect them to have typical values, although unusual market data will require unusual M and N_j . For example, if market straddle prices are unusually low then they will imply low levels of instantaneous σ , and the smaller σ is the more difficult the Black-Scholes differential problem is so solve numerically due to the dominant convection term in the partial differential equation; this is why we choose σ_{min} not to be too small.

Our approach is to generate σ using some relatively small M and N_j , but then to test the integrity of σ independently as follows. We solve the Black-Scholes straddle pricing problem, using the generated σ , using the adaptive finite element method [11]. This adaptive method solves European Black-Scholes problems to guaranteed accuracy (the level of accuracy being supplied by the user), and so we use it to find the V_{ij} (to a suitable number of significant figures) so as to establish whether or not $|e^*|$, at the minimum of $\phi + e^{**}$ subject to the constraints, is sufficiently small; this also shows whether or not the generated σ yields Black-Scholes straddle prices that match the market straddle prices to within a prescribed error tolerance. If this adaptive method shows this $|e^*|$ to be too large then we regenerate σ using slightly larger M and N_j . We continue along these lines until an appropriate choice of M and the N_j is found – with a little experience it will be easy to make an appropriate choice upfront. Thence by following this procedure we will ensure that our algorithm for generating a near-optimal instantaneous σ is robust and efficient.

Relative to the given set of market and numerical data described in Section 4.1, $M = 60$ and $(N)_j = (30 \ 40)^T$ may be considered to be an appropriate coarseness of discretisation for the minimisation since, using the adaptive method [11], we prove

$$(\mathcal{E})_{ij} = \begin{pmatrix} -2.31 & -1.07 & -0.17 & -0.29 & -0.75 \\ -6.25 & -4.65 & -3.62 & -3.96 & 1.65 \end{pmatrix}^T$$

to two-decimal accuracy. Therefore $\sum_{i,j} w_{ij} |\mathcal{E}_{ij}| < 2.65$ so that the market prices were on average replicated to within 2.65 basis points. Hence we may conclude that the bias of the volatility smile, with respect to the given set of market prices, has been effectively removed.

5 Example volatility function pricing: up-and-out call

We now illustrate how the method can be applied in pricing an exotic option. Let us price an up-and-out call option, using the generated σ of the last subsection, for example. This should serve to demonstrate the significant differences in prices and greeks that $\sigma(S, t)$ may imply when compared to a constant volatility model.

The Black-Scholes up-and-out call pricing problem is as follows: find the issue price $V(S = S(0), t = 0)$ such that

$$rV = V_t + \frac{1}{2}\sigma(S, t)^2 S^2 V_{SS} + (r - d)SV_S, \quad 0 < S < B, \quad 0 \leq t < T, \quad (5.1a)$$

$$V = (S - E)_+, \quad 0 < S < B, \quad t = T, \quad (5.1b)$$

$$V = 0, \quad S = 0, B, \quad 0 \leq t \leq T. \quad (5.1c)$$

The parameters r and d must be consistent with those used in determining σ ; it is also necessary that $T \leq T_J$. We replace the boundary condition $V = 0$ at $S = 0$ by $V = 0$ at $S = 500$; this choice well approximates the ‘theoretically correct’ one of 0 for the reasons described in Appendix A.1. We choose $B = 6500$ and $T = 1$. We solve the differential problem (5.1) accurately for each of the strike prices E_i of the last subsection using the adaptive method [11]. We depict the at-the-money issue price and payoff in Figure 6 for $S \in [4500, 6500]$.

If the up-and-out call option were priced under the assumption of constant volatility then it might be reasonable to use the implied volatility of the underlying straddle. Accordingly, we depict in Figures 7, 8 and 9 differences between the constant and functional volatility models. Here, $(V')_i$, $(V'_S)_i$ and $(V'_{SS})_i$ denote the Black-Scholes prices, deltas and gammas calculated assuming that the volatility is constant and equal to the implied volatility from the option with strike E_i and maturity $T = 1$; recall that E_3 is at-the-money. These graphs show, in particular, that pricing using the at-the-money implied volatility underestimates the price, delta and gamma compared to using the functional volatility.

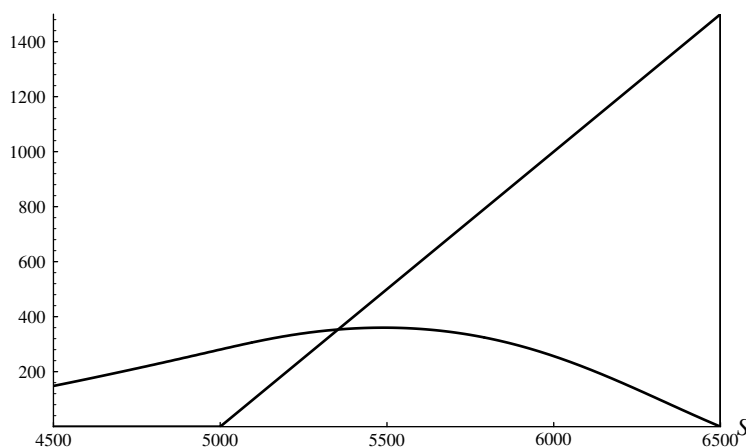


Figure 6: Up-and-out call price for $t = 0, T$.

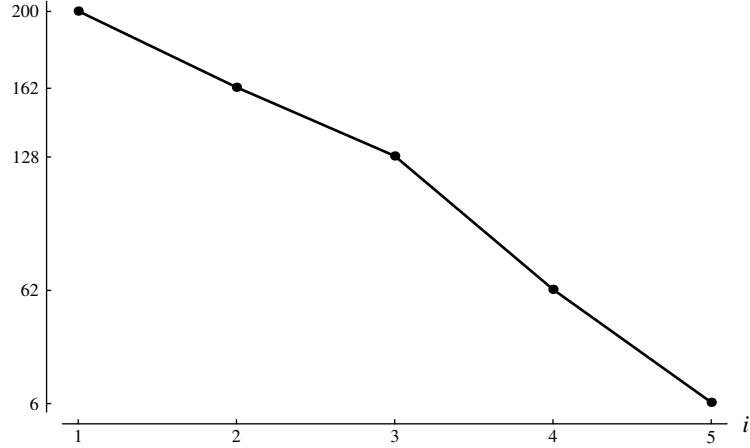


Figure 7: Basis-point differences $10^4(V - V')_i/S(0)$.

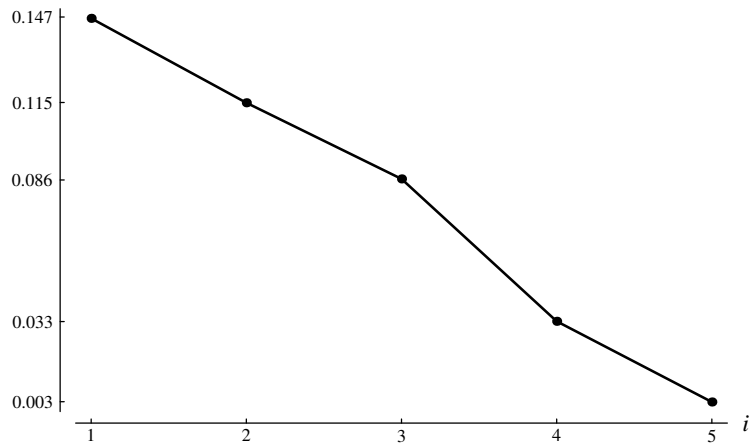


Figure 8: Delta differences $(V_S - V'_S)_i$.

Undoubtedly there will be significant differences between the rhos, vegas and thetas also, although we must consider a new way of defining vega when σ is functional – it must measure the sensitivity of the option price to the form of the functional volatility. For example, we might select a number K of appropriately perturbed market price arrays $(\bar{V})_{ijk}$, for $k = 1, \dots, K$, then generate $\sigma_k(S, t)$ for each of these, and then take vega as some relative measure of the difference between the option prices implied by ‘today’s functional σ ’ and each of the σ_k : for a given strike and maturity, we might take

$$\text{vega} = \left((P + 1)(Q + 1) \sum_k \frac{|V(\sigma) - V(\sigma_k)|^2}{\sum_{p=0}^P \sum_{q=0}^Q |\Sigma_{pq} - \Sigma_{pqk}|^2} \right)^{1/2}.$$

6 Conclusions

We have proposed a method of generating a functional form for σ that minimises an appropriate measure of the difference between Black-Scholes vanilla prices and a given

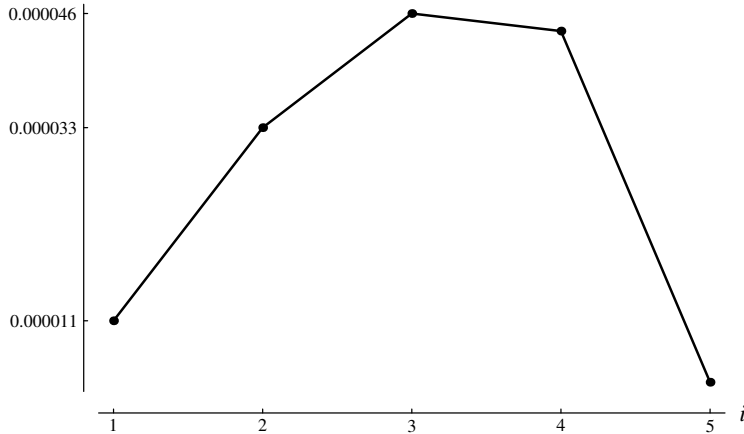


Figure 9: Gamma differences $(V_{SS} - V'_{SS})_i$.

set of market vanilla prices. Because this instantaneous σ is chosen as a spline, it is continuously well-defined in space-time. Moreover, because the optimised strategy is regularised and integrity tested, σ should not only be the smoothest such spline but it is proved to replicate the given set of market prices to within a prescribed error tolerance. Therefore, with respect to this given set of market prices, it effectively removes the bias of the volatility smile exhibited by the implied volatility model.

The computational method is not overly time-consuming but will depend on the number of given market prices. Therefore attempting to minimise the difference between Black-Scholes vanilla prices and an overly large set of market vanilla prices is inappropriate. Moreover, the complete set of yesterday's (for example) market prices do not necessarily accurately reflect the biases of the Black-Scholes model for many months to come. Indeed, there is clearly a trade-off between trying to match all market prices and trying to predict some (and hence spot arbitrages). It is clear that 'outlying' market prices, whether mispricings or misquotations, may significantly skew the volatility surface, so there is a case for a pre-processing procedure that identifies these prices and rejects them before the minimisation. Also, we need not match all of the market prices of options that are far from the money since they may increase the computational complexity of the model considerably while only have a small effect on the prices and greeks of exotic options that are calculated using the σ .

Having generated σ for a given underlying it may be used consistently in the pricing of exotic options and may be employed by a variety of lattice, Monte Carlo and partial differential equation pricing methods. We note that it is particularly advantageous that σ is a simple function, defined for a large range of spot price, that is particularly smooth in nature and is proved to replicate a given set of market prices to sufficient accuracy.

We end with a question that can only be answered by careful empirical work: despite the advantages of the functional volatility model that we have outlined, is the model more effective in pricing and hedging exotic options than either the simple remedy of using an array of constant implied volatilities or more complex models such as stochastic volatility?

References

- [1] Avellaneda, M., Friedman, C., Holmes, R. and Samperi, D. Calibrating volatility surfaces via relative entropy minimization. *Applied Mathematical Finance*, 4:37–64, 1997.
- [2] Barle, S. and Cakici, N. Growing a smiling tree. *Risk Magazine*, 9(10):76–81, 1995.
- [3] Black, F. and Scholes, M. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(1):637–654, 1973.
- [4] Cox, J.C. and Ross, S.A. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:145–166, March 1976.
- [5] Derman, E. and Kani, I. Riding on a smile. *Risk Magazine*, 7(2):32–39, 1994.
- [6] Dupire, B. Pricing with a smile. *Risk Magazine*, 7(1):18–20, 1994.
- [7] Gill, P.E. and Murray, W. Minimization subject to bounds on the variables. Technical Report NAC 72, National Physical Laboratory Report, 1976.
- [8] Heston, S.L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.
- [9] Hull, J. and White, A. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2):281–300, 1987.
- [10] Hull, J.C. *Options, Futures, and other Derivatives - third edition*. Prentice-Hall, 1997.
- [11] Jackson, N. and Süli, E. Adaptive finite element solution of 1D European option pricing problems. Technical Report 97/05, Oxford University Computing Laboratory, 1997.
- [12] Jackwerth, J.C. Implied binomial trees: Generalizations and empirical tests. Technical report, Berkeley, 1996.
- [13] Jackwerth, J.C. and Rubinstein, M. Recovering probability distributions from option prices. *Journal of Finance*, 51:1611–1631, 1996.
- [14] Lagnado, R. and Osher, S. A technique for calibrating derivative security pricing models: numerical solution of an inverse problem. *Journal of Computational Finance*, 1(1):13–25, 1997.
- [15] Merton, R.C. Option pricing when underlying stock returns are discontinuous. *The Journal of Financial Economics*, 3:125–144, 1976.
- [16] Rubinstein, M. Implied binomial trees. *Journal of Finance*, 69:771–818, 1994.
- [17] Schwarz, H.R. *Numerical Analysis - a comprehensive introduction*. John Wiley & Sons, 1989.
- [18] Tykhonov, M. Regularization of incorrectly posed problems. *Soviet Mathematics*, 4:1624–1627, 1963.
- [19] Wilmott, P., Dewynne, J.N. and Howison, S. *Option Pricing: mathematical models and computation*. Oxford Financial Press, 1993.

Appendix A Black-Scholes straddle prices

A.1 Differential problems

To evaluate the component V_{ij} of F , we must solve for each pair (E_i, T_j) the following Black-Scholes straddle pricing problem: find the issue price $V(S = S(0), t = 0)$ such that

$$r_j V = V_t + \frac{1}{2} \sigma(S, t)^2 S^2 V_{SS} + (r_j - d_j) S V_S, \quad S_{min} < S < S_{max}, \quad 0 \leq t < T_j \quad \text{(A.1a)}$$

$$V = |S - E_i|, \quad S_{min} < S < S_{max}, \quad t = T_j, \quad \text{(A.1b)}$$

$$V = E_i e^{-r_j(T_j-t)}, \quad S = S_{min}, \quad 0 \leq t \leq T_j, \quad \text{(A.1c)}$$

$$V = S_{max} e^{-d_j(T_j-t)} - E_i e^{-r_j(T_j-t)}, \quad S = S_{max}, \quad 0 \leq t \leq T_j. \quad \text{(A.1d)}$$

Note that in contrast to the ‘theoretically correct’ differential problem for vanilla options (see [19, pages 425-426]) in which S_{min} should be chosen to be 0 and S_{max} should be chosen to denote ∞ (in accordance with the consideration that vanilla option value can only be deterministic in the limits $S \rightarrow 0, \infty$) we make the restriction that $0 < S_{min} < S_{max} < \infty$ where S_{min} and S_{max} are simply chosen to be sufficiently small and large respectively. We easily find (by solving the differential problem for the difference between the restricted and ‘theoretically correct’ differential problems) that the error incurred by the restriction is extremely small. For example, by simply choosing $S_{min} = S(0)/10$ and $S_{max} = 10S(0)$ we find for straddle options for realistic financial parameter choices that the incurred error is several orders of magnitude less than the modelling error within trading neighbourhoods of the spot price. We make the restrictions on S_{min} and S_{max} , as is possibly clear, a) because it is difficult to solve numerically on a infinite interval and b) so as to ensure that the problem always remains non-degenerate parabolic.

It is of clearly of paramount importance that the differential problem (A.1) involved in the minimisation is solved to sufficient accuracy and in the least possible time. If its space-time discretisation is either too coarse or too fine then the generated instantaneous σ will either be inappropriate or will have taken too long to generate. Financial time constraints are critical inasmuch as it may be necessary to generate several instantaneous σ for the same underlying (for a scenario analysis) or for different underlyings, and on a regular basis.

We solve the problem (A.1) as follows: we first transform the financial variables and we then solve the resulting transformed problem by a piecewise quadratic finite element method in space and a Crank-Nicolson finite difference method in time. We perform the transformation, which we describe below, for reasons of efficiency; the Péclet number of the resulting partial differential equation is maximised near to the financial region of interest which should improve the performance of the finite element method. We find experimentally that the piecewise quadratic finite element method in space, which we describe in the next subsection, is considerably faster than a variety of standard finite difference methods in space (for example those described in [19]) relative to the level of accuracy we require.

We transform the financial variables by

$$S = S(0)e^x, \quad t = T_j - \tau, \quad V = S(0)e^{k_1 x + k_2 \tau} u,$$

having defined the constants

$$k_1 := 1/2 - (r_j - d_j)/\bar{\sigma}^2, \quad k_2 := -r_j - k_1^2 \bar{\sigma}^2/2,$$

where $\bar{\sigma}$ is a constant derived from σ by some form of simple averaging (for example, we might choose $\bar{\sigma}$ as the at-the-money implied volatility relative to some T_j), to obtain the following transformed problem: find $u(x=0, \tau=T_j)$ such that

$$u_\tau = \alpha u_{xx} + \beta u_x + \gamma u, \quad a < x < b, \quad 0 \leq \tau < T_j, \quad (\text{A.2a})$$

$$u = u_0, \quad a < x < b, \quad \tau = 0, \quad (\text{A.2b})$$

$$u = u_a, \quad x = a, \quad 0 \leq \tau \leq T_j, \quad (\text{A.2c})$$

$$u = u_b, \quad x = b, \quad 0 \leq \tau \leq T_j, \quad (\text{A.2d})$$

having defined $a := \log(S_{min}/S(0))$, $b := \log(S_{max}/S(0))$, $c := \log(E_i/S(0))$ and the functions

$$\alpha := \sigma^2/2,$$

$$\beta := r_j - d_j + \alpha(2k_1 - 1),$$

$$\gamma := -r_j - k_2 + k_1(\beta - \alpha k_1),$$

$$u_0 := |e^{(1-k_1)x} - e^{-k_1x+c}|,$$

$$u_a := e^{-k_1a+c-(r_j+k_2)\tau},$$

$$u_b := e^{(1-k_1)b-(d_j+k_2)\tau} - e^{-k_1b+c-(r_j+k_2)\tau}.$$

A.2 Discretisation method

To solve the problem (A.2) (numerically by our piecewise quadratic finite element method in space) we proceed by considering its weak formulation: find $u \in H_E^1$ such that, for each $\tau > 0$, and for all functions $\bar{v} \in H_{E_0}^1$,

$$(u_\tau, \bar{v}) + (\alpha u_x, \bar{v}_x) + ((\alpha_x - \beta)u_x - \gamma u, \bar{v}) = 0; \quad (\text{A.3})$$

recall that by definition

$$H_E^1 := \{w \in L^2(a, b) \mid w_x \in L^2(a, b), \quad w(a) = u_a, \quad w(b) = u_b\}$$

while $H_{E_0}^1$ is similarly defined except with respect to homogeneous boundary conditions, and recall the inner product notation

$$(f, g) \equiv \int_a^b fg \, dx$$

for some f and g . Then, given the M -dimensional trial-subspace $S_E^h \subset H_E^1$, the *continuous Galerkin principle* is to find the approximation $U(x, \tau)$ to $u(x, \tau)$ that (for each $\tau > 0$) lies in S_E^h and satisfies

$$(U_\tau, v) + (\alpha U_x, v_x) + ((\alpha_x - \beta)U_x - \gamma U, v) = 0 \quad (\text{A.4})$$

for every $v \in S_{E_0}^h$; we choose $U_0 \equiv U|_{\tau=0}$ to be the piecewise quadratic interpolant of u_0 . We define S_E^h as the space of C^0 piecewise quadratic functions, satisfying the (essential) Dirichlet boundary conditions, on the spatial mesh

$$a = x_0 < x_1 < \dots < x_i < \dots < x_{2M} = b.$$

If ξ denotes a local co-ordinate relative to the interval $h_i := x_{2i} - x_{2i-2}$ where

$$x = \frac{1}{2}(1 - \xi)x_{2i-2} + \frac{1}{2}(1 + \xi)x_{2i}$$

for the i th quadratic element (for $i = 1, 2, \dots, M$) then the basis functions are

$$N_1 := -\frac{1}{2}\xi(1 - \xi), \quad N_2 := 1 - \xi^2, \quad N_3 := \frac{1}{2}\xi(1 + \xi).$$

Then we have, for the i th quadratic element,

$$U(x(\xi), \tau) = U_{2i-2}(\tau)N_1 + U_{2i-1}(\tau)N_2 + U_{2i}(\tau)N_3.$$

The contributions I_i to the i th integrals of the approximated weak formulation (A.4) may be written

$$I_i = \begin{pmatrix} v_{2i-2}(\tau) \\ v_{2i-1}(\tau) \\ v_{2i}(\tau) \end{pmatrix} \cdot \left(M_i \begin{pmatrix} U_{\tau_{2i-2}}(\tau) \\ U_{\tau_{2i-1}}(\tau) \\ U_{\tau_{2i}}(\tau) \end{pmatrix} + K_i \begin{pmatrix} U_{2i-2}(\tau) \\ U_{2i-1}(\tau) \\ U_{2i}(\tau) \end{pmatrix} \right),$$

whose components are defined as follows. The 3×3 element mass matrix M_i has entries

$$M_{ipq} = \frac{h_i}{2} \int_{-1}^1 N_p N_q d\xi,$$

so that

$$M_i = \frac{h_i}{30} \begin{pmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4 \end{pmatrix},$$

and the 3×3 element stiffness matrix K_i has entries

$$K_{ipq} = \int_{-1}^1 \left(\frac{2}{h_i} \alpha N_p \xi N_q \xi + (\alpha_x - \beta) N_p N_q \xi - \frac{h_i}{2} \gamma N_p N_q \right) d\xi;$$

it is most appropriate to approximate these integrals using the two-point Gaussian quadrature rule (see [17, pages 367-368] for example). Summing the contributions I_i over i , noting that it is enough to apply the continuous Galerkin principle only to the spatial basis of \underline{v} , yields the full system of equations

$$M^f \underline{U}_\tau^f + K^f \underline{U}^f = 0,$$

where

$$\underline{U}^f = (U_0(\tau) \quad U_1(\tau) \quad \cdots \quad U_{2M}(\tau))^T,$$

and with the pentadiagonal full mass matrix

$$M^f = \frac{1}{30} \begin{pmatrix} 4h_1 & 2h_1 & -h_1 & & & \cdots & 0 \\ 2h_1 & 16h_1 & 2h_1 & 0 & & & \vdots \\ -h_1 & 2h_1 & 4h_1 + 4h_2 & 2h_2 & -h_2 & & \\ & 0 & 2h_2 & 16h_2 & 2h_2 & 0 & \\ & & -h_2 & 2h_2 & 4h_2 + 4h_3 & 2h_3 & -h_3 \\ & & & 0 & 2h_3 & 16h_3 & \ddots & \ddots \\ & & & & -h_3 & \ddots & \ddots & 2h_M & -h_M \\ \vdots & & & & & \ddots & 2h_M & 16h_M & 2h_M \\ 0 & \cdots & & & & & -h_M & 2h_M & 4h_M \end{pmatrix}.$$

Then application of the Dirichlet boundary conditions yields the global system of equations

$$M\underline{U}_\tau + K\underline{U} = \underline{bc}^*, \quad (\text{A.5})$$

where

$$\underline{U} = (U_1(\tau) \quad U_2(\tau) \quad \cdots \quad U_{2M-1}(\tau))^T, \\ \underline{bc}^* = \begin{pmatrix} -2h_1 u_{a_\tau}(\tau) - K_{1,0} u_a(\tau) \\ h_1 u_{a_\tau}(\tau) - K_{2,0} u_a(\tau) \\ 0 \\ \vdots \\ 0 \\ h_M u_{b_\tau}(\tau) - K_{2M-2,2M} u_b(\tau) \\ -2h_M u_{b_\tau}(\tau) - K_{2M-1,2M} u_b(\tau) \end{pmatrix},$$

the global matrices being simply the full matrices less their first and last rows and columns.

We discretise in time via a Crank-Nicolson type finite difference method. Specifically, on the temporal mesh

$$0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots < \tau_{N_j} = T_j,$$

defining $k_n := \tau_n - \tau_{n-1}$, we average first-order accurate forward and backward temporal difference approximations to (A.5). Thence we obtain the following *matrix problem*: find

$$\underline{U}^n = (U_1^n \quad U_2^n \quad \cdots \quad U_{2M-1}^n)^T,$$

for $n = 1, 2, \dots, N_j$ in succession, such that

$$(M + \frac{1}{2}k_n K^n) \underline{U}^n = (M - \frac{1}{2}k_n K^{n-1}) \underline{U}^{n-1} + \underline{bc}, \quad (\text{A.6})$$

where

$$\underline{U}^0 = (u_{0_1} \quad u_{0_2} \quad \cdots \quad u_{0_{2M-1}})^T, \\ \underline{bc} = \begin{pmatrix} -2h_1(u_a^n - u_a^{n-1}) - \frac{1}{2}k_n(K_{1,0}^n u_a^n + K_{1,0}^{n-1} u_a^{n-1}) \\ h_1(u_a^n - u_a^{n-1}) - \frac{1}{2}k_n(K_{2,0}^n u_a^n + K_{2,0}^{n-1} u_a^{n-1}) \\ 0 \\ \vdots \\ 0 \\ h_M(u_b^n - u_b^{n-1}) - \frac{1}{2}k_n(K_{2M-2,2M}^n u_b^n + K_{2M-2,2M}^{n-1} u_b^{n-1}) \\ -2h_M(u_b^n - u_b^{n-1}) - \frac{1}{2}k_n(K_{2M-1,2M}^n u_b^n + K_{2M-1,2M}^{n-1} u_b^{n-1}) \end{pmatrix}.$$

Thus far we have not specified the location of x_i on (a, b) and τ_n on $(0, T_j)$. We might specify them by prescribing a uniform space-time discretisation by choosing $h_i = (b - a)/M$ ($i = 1, \dots, M$) and $k_n = T_j/N_j$ ($n = 1, \dots, N_j$) but it is much more efficient to prescribe non-uniform sets of h_i and k_n that account for the nature of the initial and boundary data. Therefore, since we are solving a transformed straddle pricing problem, we have chosen the h_i so that the spatial mesh is concentrated finely near to the spot price and we have chosen the k_n so that the temporal mesh is concentrated finely near to $\tau = 0$; it is important that any such meshes are smoothly

varying so as to prevent ill-conditioning in the matrices of (A.6).

Thus we see that the transformed differential problem (A.2) has been discretised in space by approximating its solution space by a space of M -dimensional piecewise quadratic functions while it has been discretised in time by finite differencing over N_j time-levels. Therefore to obtain the approximation to the straddle price $V(S = S(0), t = 0)$ for each pair (E_i, T_j) , we must first solve the pentadiagonal matrix problem (A.6) (via band Gauss elimination, say), so as to obtain the piecewise quadratic U^{N_j} , and then we must relate back to financial variables by the formula

$$V(S = S(0), t = 0) \approx S(0)e^{k_2 T_j} U^{N_j}.$$

Lastly, we remark that considerable care should be taken when programming the above procedure. Significant efficiency savings can be made by noting, for example, that the stiffness matrix K^n need only be constructed once for each maturity, i.e. once for the pairs $(E_1, T_j), \dots, (E_I, T_j)$ for each j .