The equity option volatility smile: an implicit finite-difference approach

Leif B. G. Andersen and Rupert Brotherton-Ratcliffe

This paper illustrates how to construct an unconditionally stable finite-difference lattice consistent with the equity option volatility smile. In particular, the paper shows how to extend the method of forward induction on Arrow–Debreu securities to generate local instantaneous volatilities in implicit and semi-implicit (Crank–Nicholson) lattices. The technique developed in the paper provides a highly accurate fit to the entire volatility smile and offers excellent convergence properties and high flexibility of asset- and time-space partitioning. Contrary to standard algorithms based on binomial trees, our approach is well suited to price options with discontinuous payouts (e.g. knock-out and barrier options) and does not suffer from problems arising from negative branching probabilities.

1. INTRODUCTION

The Black–Scholes option pricing formula (Black and Scholes 1973, Merton 1973) expresses the value of a European call option on a stock in terms of seven parameters: current time t, current stock price S_t, option maturity T, strike K, interest rate r, dividend rate γ, and volatility[1] θ. As the Black–Scholes formula is based on an assumption of stock prices following geometric Brownian motion with constant process parameters, the parameters r, γ, and θ are all considered constants independent of the particular terms of the option contract. Of the seven parameters in the Black–Scholes formula, all but the volatility θ are, in principle, directly observable in the financial market. The volatility θ can be estimated from historical data or, as is more common, by numerically inverting the Black–Scholes formula to back out the level of θ—the *implied volatility*—that is consistent with observed market prices of European options.

Although the Black–Scholes pricing formula has become the *de facto* standard in many options markets, it is generally recognized that the assumptions underlying the formula are imperfect. For example, the existence of term structures in interest rates and dividends indicate that r and γ are not constants but (at least) functions of t and T. More seriously, backing out implied volatilities from the Black–Scholes formula frequently yields θ's that are functions of maturity and stroke. Dependent on the shape of the mapping $K \mapsto \theta(S_t, t; K, T)$, the phenomenon of time- and strike-dependent volatilities is referred to as the *volatility smile* or the *volatility skew*; its existence indicates that the true probability distribution of stock prices deviates from the ideal log-normal distribution of the a Black–Scholes analysis. In the 70s and early 80s, the volatility smile in US equity options was relatively mild and frequently either ignored by market participants or handled in an *ad hoc* manner; indeed, using S&P 500 options data from 1976 to 1978, Rubinstein (1985) detects no economic significance of the errors associated with using a constant volatility for options with the same maturity but different strikes. The crash in 1987, however, appears to have increased the likelihood assigned by the financial markets to extreme stock market movements, in particular large *downward* movements. Sometimes known as 'crash-o-phobia', this change in view of stock price dynamics has resulted in a persistent, pronounced volatility smile in current options markets (Shimko 1993, Rubinstein 1994).

Traditionally, the problems of nonconstant parameters in the models have been handled pragmatically by simply maintaining vectors and tables of r, γ, and θ to be used with different option maturities and strikes. Although this approach—by construction—works well for European options, it is unsuited for pricing of more complicated structures such as exotic options and options with early exercise features (Bermuda and American options). Consider, for example, a 2-year knock-out option with a strike of $100 and a knock-out level of $90. In interpolating a value of θ for the knock-out option from a (K, T) table of implied call option volatilities, should one use $100 or $90 (or some third value) for K? And should one use 2 years as T or, given that the option can be knocked out before it reaches its final maturity, some lower value?[2]

To answer questions like the one above, many researchers have attempted to develop models that are consistent with the existence of a volatility smile. One line of research has focused on enriching the Black–Scholes analysis by introducing additional sources of risk, including Poisson jumps (Merton 1976) and stochastic volatility (Hull and White 1987). Besides being difficult to implement and calibrate, such models lack completeness and do not allow for arbitrage-free pricing. To preserve completeness and avoid having to make assumptions about investor preferences and behavior, many newer approaches stay within the Black–Scholes one-factor diffusion framework, but introduce extra degrees of freedom by allowing the instantaneous local volatility to be a function of both time and stock levels. As it turns out, this framework is sufficiently rich to allow a perfect fit to most reasonable volatility smiles and at the same time preserves completeness and allows for application of the usual arbitrage-free pricing techniques.

The option models based on one-factor stock diffusions take several forms. In one approach, the local volatility function is prescribed directly, typically as a well-behaved function of only a few parameters (Cox and Ross 1976, Beckers 1980, Platen and Schweizer 1994). The specification of the volatility function can for example be based on a microeconomic analysis of interactions between agents in the options market, as in Platen and Schweizer (1994). Although sometimes quite realistic smiles and skews can be generated from a direct parametrization of local volatility, it is, in general, not likely that this approach will lead to a satisfactory fit to the market smile. In this paper, we instead choose to focus on an alternative, more recent, modeling technique which takes the market volatility smile as a direct input and, through numerical or analytical techniques, backs out an implied local volatility function that is consistent with the observed volatility smile. One early effort along these lines was made by Dupire (1994), who develops a continuous-time theory in a setting without interest rates and dividends. Dupire's continuous-time results have been supplemented by a number of discrete-time numerical methods, mostly set in a binomial framework. The fit to the volatility smile is obtained through careful manipulation of the local branching probabilities in the binomial tree. Examples of such so-called *implied binomial trees* can be found in Rubinstein (1994, 1995), Derman and Kani (1994), Barle and Cakici (1995), and Chriss (1996).

The method of implied binomial trees offers a relatively straightforward approach to fitting the volatility smile, but suffers from a number of fundamental problems. First, the degrees of freedom at each tree node are not sufficiently high to guarantee that all binomial branching probabilities are nonnegative,[3] particularly in environments with high interest rates and steep volatility smiles. The heuristic rules that are typically applied to override nodes where illegal branching occurs (see Derman and Kani 1994) are not only unsatisfactory but result in loss of local process information that can easily compound up to significant pricing errors (Barle and Cakici 1995). A second problem of binomial trees has been documented by Boyle and Lau (1994), who illustrate how using binomial trees to price options with discontinuous payouts (such as barrier and knock-out options) can lead to extremely erratic convergence behavior unless care is taken to align the asset partitioning of the tree with the option barrier. As binomial trees have very limited flexibility in setting the partitioning of the asset space—in fact, the asset grid can essentially only be affected indirectly through the choice of number of time-steps—this alignment process can frequently put severe constraints on the overall design of the lattice. For implied binomial trees, the alignment process is generally not even possible, as the time- and asset-varying nature of the branching process results in trees where the asset-partitioning of each time slice is unique and not aligned with the asset levels of other slices.

Whereas the implied binomial tree is primarily based on a discretization of the stock price process, this paper will focus on developing a discrete-time model by discretizing the fundamental no-arbitrage partial differential equation (PDE). This discretization is accomplished by an adaptation of the method of *finite differences* (see, for example, Brennan and Schwartz 1978, Courtadon 1982, Geske and Shastri 1985, Hull and White 1990, and Dewynne *et al*. 1993). The application of one particularly simple finite-difference scheme, the so-called *explicit scheme* (or *trinomial tree*), to the volatility smile problem has been described by Dupire (1994) and, in a purely probabilistic setting, by Derman *et al*. (1996). As we will show in the paper, the explicit finite-difference method, however, suffers from many of the same problems as the binomial tree and is prone to instability. In this paper, we instead focus on an alternative class of algorithms known as implicit and semi-implicit (Crank–Nicholson) finite-difference schemes. While somewhat more complicated to evaluate and calibrate, the implicit and semi-implicit schemes are shown to exhibit much better stability and convergence properties than trinomial and binomial trees. Further, contrary to the binomial algorithm, the algorithms developed in this paper do not involve explicit adjustments of branching probabilities and allow for completely independent prescription of the stock- and time-partitionings. The high partitioning flexibility permits control of convergence behavior as it allows for perfect alignment of time- and asset-slices with important dates (e.g. dividends, average sampling dates, trigger observation dates, etc.) and price levels (e.g. strikes, barriers, etc.).

While our numerical approach is different, our paper is similar in spirit to the original work by Dupire (1994). In particular, we assume the existence of a complete, spanning set of European call option prices, which, in practice, requires usage of extrapolation and interpolation methods. An alternative approach (e.g. Avallaneda *et al*. 1996, Lagnado and Osher 1997, and Brown and Toft 1996) is to work only with actively traded options and 'fill in' the gaps indirectly through assumptions about market behavior and regularity. While this approach has its merits, it yields less control over the resulting volatility surfaces and, as large-scale nonlinear optimization is typically necessary, is much slower than the method used in this paper.

The rest of this paper is organized as follows. In Section 2, we summarize the continuous-time theory of Dupire (1994) and provide some extensions to include nonzero dividends and interest rates. Section 3, the main section of the paper, develops the theory of our implicit finite-difference approach. In Section 4, we test the accuracy and convergence properties of the finite-difference algorithm and exemplify its application to exotic options by pricing down-and-out knock-out call options. Finally, Section 5 summarizes the results of the paper and briefly discusses extensions and generalizations.

2. CONTINUOUS TIME

In this section, we present the continuous-time theory behind the one-factor diffusion approach to modeling the dynamics of the volatility smile. The material in this section is based on Dupire (1994), but is set in a more general framework.

Let us consider a frictionless economy in which a traded asset S is driven by a one-factor diffusion process of the form

$$\mathrm{d}S_t/S_t = \mu(S_t, t)\,\mathrm{d}t + \sigma(S_t, t)\,\mathrm{d}W_t, \qquad S_0 = S_{\mathrm{ini}} > 0, \quad t \in [0, \tau]. \tag{1}$$

for some fixed trading horizon τ and some positive constant time 0 value S_{ini}. In (1), W_t is a Brownian motion with respect to the real-world probability measure and $\mu,\sigma : \mathbb{R}^+ \times [0,\tau] \to \mathbb{R}$ are deterministic functions sufficiently well behaved to ensure that (1) has a unique solution (see Arnold 1974: Chap. 6). We will assume that S pays dividends at a time-varying, but deterministic, rate of $\gamma(t)$. For fixed $t \in [0,\tau]$ and all or, written in terms of the observed implied volatility smile[4] $\theta(S_t, t; K, T)$,

$$\sigma^2(K,T) = \frac{2\dfrac{\partial\theta}{\partial T} + \dfrac{\theta}{T-t} + 2K[r(T)-\gamma(T)]\dfrac{\partial\theta}{\partial K}}{K^2\left[\dfrac{\partial^2\theta}{\partial K^2} - d_+\sqrt{T-t}\left(\dfrac{\partial\theta}{\partial K}\right)^2 + \dfrac{1}{\theta}\left(\dfrac{1}{K\sqrt{T-t}} + d_+\dfrac{\partial\theta}{\partial K}\right)^2\right]}, \qquad (16)$$

where d_+ is defined in (9).

Proof. Equation (15) follows immediately from (14), and (16) follows, after some manipulations, from (15) and (9). To verify that $\sigma(K, T)$ is a real number, i.e. that $\sigma^2(K, T) \geqslant 0$, we notice from (11) that it suffices to show that the numerator in (15) is nonnegative in the absence of arbitrage. Portfolio dominance arguments similar to those in Merton (1973) imply the following result:

$$e^{\int_T^{T_1}\gamma(u)\,du} C\left(S_t, t; K\, e^{\int_T^{T_1}[r(u)-\gamma(u)]\,du}, T_1\right) \geqslant C(S_t, t; K, T), \qquad T_1 > T.$$

Setting $T_1 = T + \varepsilon$ in the left-hand side of the above inequality and evaluating the limit as $\varepsilon \to 0_+$ yields

$$\frac{\partial C}{\partial T} + \gamma(T)C + K([r(T) - \gamma(T)]\frac{\partial C}{\partial K} \geqslant 0. \qquad \square$$

For the special case of strike-independent implied volatility, (16) reduces to the well-known expression

$$\sigma^2(T) = \theta^2(T) + 2(T-t)\theta(T)\frac{\partial\theta}{\partial T},$$

or

$$\frac{1}{T-t}\int_t^T \sigma^2(s)\,ds = \theta^2(T).$$

3. DISCRETE TIME

While equations (15) and (16) in combination with the no-arbitrage PDE (4) exhaust the theoretical specification of the volatility smile model, in practice numerical methods must be introduced to calculate the prices of specific contingent claims. As discussed in Section 1, most such schemes suggested in the current literature are based on a binomial approximation of the stochastic differential equation (1). In this section, we will develop an alternative to the binomial method using the method of finite differences.

3.1 Discretization Scheme

To increase the efficiency of the finite-difference discretization, we first shift variables in the PDE (4). Specifically, we put $x = \ln S$ and $H(x, t) = V(S, t)$ so that the governing equation becomes

$$\frac{\partial H(x,t)}{\partial t} + \tfrac{1}{2} v(x,t) \frac{\partial^2 H(x,t)}{\partial x^2} + b(x,t) \frac{\partial H(x,t)}{\partial x} = r(t) H(x,t), \tag{17}$$

where

$$b(x,t) = r(t) - \gamma(t) - \tfrac{1}{2} v(x,t), \qquad v(x,t) = \sigma^2(S,t) = \sigma^2(\mathrm{e}^x, t).$$

At this point, we could use the continuous-time dividend term structure $\gamma(t)$, the interest rate term structure (3) and the instantaneous volatility equations (15) and (16) to discretize (17) directly. However, as the coefficients in (17) would then all be based on results from a continuous-time setting, such a discretization would only in the limit yield correct prices of traded bonds and stock derivatives. To improve convergence and accuracy of discrete-time prices, we replace the continuous-time coefficients in (17) by unknown functions $\hat{r}(t)$, $\hat{b}(x,t)$, and $\hat{v}(x,t)$ which shall be solved for so that our discretization of the backward PDE will return the correct market prices of stock forwards, zero-coupon bonds, and European options. We point out that $\hat{r}(t)$, $\hat{b}(x,t)$, and $\hat{v}(x,t)$ will depend on both the discretization scheme and the selected spacing between grid points.

Now consider determining the time-0 value of a contingent claim with final maturity $0 < T < \tau$. To discretize (17), we divide the (x,t) plane into a uniformly spaced mesh with $M+2$ nodes along the t axis and $N+2$ nodes along the x axis:

$$x_i = x_0 + i\Delta_x = x_0 + i\frac{x_{N+1} - x_0}{N+1}, \qquad i = 0, \dots, N+1, \tag{18a}$$

$$t_j = j\Delta_t = j\frac{T}{M+1}, \qquad j = 0, \dots, M+1. \tag{18b}$$

The indices $i = 0$, $i = N+1$, $j = 0$, $j = M+1$ signify the limits of the mesh for which boundary conditions must be prescribed. The values of x_0 and x_{N+1} should be set sufficiently low and high, respectively, to ensure that most of the statistically significant x space is captured by the mesh.[5] Without loss of generality, we assume that the time-0 stock value S_{ini} is contained in the mesh,[6] i.e.

$$x_{\mathrm{ini}} \equiv \ln S_{\mathrm{ini}} = x_\beta, \tag{19}$$

for some integer $\beta \in [1, N]$. We point out that the finite-difference method does not rely on an equidistant partitioning of t and x space (as in (18a,b)); for the sake of simplicity, however, we maintain the assumption of a uniform mesh throughout this paper.

At an arbitrary node (x_i, t_j), with $i = 1, \dots, N$, $j = 0, \dots, M$ in the grid (18a,b) we introduce the following difference approximations to the terms in the PDE (17):

$$\frac{\partial H}{\partial t} \approx \frac{H(x_i, t_{j+1}) - H(x_i, t_j)}{\Delta_t}, \tag{20a}$$

$$\frac{\partial H}{\partial x} \approx (1 - \Theta)\frac{H(x_{i+1}, t_j) - H(x_{i-1}, t_j)}{2\Delta_x} + \Theta\frac{H(x_{i+1}, t_{j+1}) - H(x_{i-1}, t_{j+1})}{2\Delta_x}, \tag{20b}$$

$$\frac{\partial^2 H}{\partial x^2} \approx (1-\Theta)\frac{H(x_{i+1},t_j)-2H(x_i,t_j)+H(x_{i-1},t_j)}{\Delta_x^2} + \Theta\frac{H(x_{i+1},t_{j+1})-2H(x_i,t_{j+1})+H(x_{i-1},t_{j+1})}{\Delta_x^2}. \quad (20c)$$

The parameter $\Theta \in [0,1]$ determines the time at which partial derivatives w.r.t. x are evaluated. If $\Theta = 0$, the x derivatives are evaluated at time t_j and the differencing scheme gives rise to the *fully implicit finite-difference method*. If $\Theta = 1$, the x derivatives are evaluated one time-step ahead, at t_{j+1}, and the resulting scheme is known as the *explicit finite-difference method*. Finally, when $\Theta = \frac{1}{2}$, the x derivatives are evaluated half a time-step ahead, at $\frac{1}{2}(t_j + t_{j+1})$; the resulting scheme is an average of the explicit and implicit schemes known as the *Crank–Nicholson* scheme. Values of Θ different from 0, $\frac{1}{2}$, and 1 are possible but little used in practice.

Plugging (20a–c) into (17) and substituting $\hat{b}$, $\hat{r}$, $\hat{v}$ for b, r, and v, respectively, yields the recursive relation for $i = 1, \ldots, N$ and $j = 0, \ldots, M$ (with $H_{i,j} \equiv H(x_i, t_j)$, etc.):

$$\begin{aligned} &H_{i-1,j}\big(-\tfrac{1}{2}\alpha(1-\Theta)(\hat{v}_{i,j} - \Delta_x \hat{b}_{i,j})\big) \\ &\quad + H_{i,j}\big(1 + \hat{r}_j\Delta_t + \alpha(1-\Theta)\hat{v}_{i,j}\big) + H_{i+1,j}\big(-\tfrac{1}{2}\alpha(1-\Theta)(\hat{v}_{i,j} + \Delta_x \hat{b}_{i,j})\big) \\ &= H_{i-1,j+1}\big(\tfrac{1}{2}\alpha\Theta(\hat{v}_{i,j} - \Delta_x \hat{b}_{i,j})\big) + H_{i,j+1}(1 - \alpha\Theta\hat{v}_{i,j}) + H_{i+1,j+1}\big(\tfrac{1}{2}\alpha\Theta(\hat{v}_{i,j} + \Delta_x \hat{b}_{i,j})\big), \quad (21) \end{aligned}$$

where $\alpha \equiv \Delta_t/\Delta_x^2$.

Equation (21) can be written compactly in matrix notation as

$$[(1 + \hat{r}_j\Delta_t)\boldsymbol{I} - (1-\Theta)\boldsymbol{M}_j]\boldsymbol{H}_j = \big(\Theta\boldsymbol{M}_j + \boldsymbol{I}\big)\boldsymbol{H}_{j+1} + \boldsymbol{B}_j, \qquad j = 0, \ldots, M, \qquad (22)$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix, $\boldsymbol{H}_j$ and $\boldsymbol{H}_{j+1}$ are $N \times 1$ vector of contingent claim values,

$$\boldsymbol{H}_j \equiv \begin{bmatrix} H_{1,j} \\ H_{2,j} \\ \vdots \\ H_{N,j} \end{bmatrix},$$

$\boldsymbol{B}_j$ is a $N \times 1$ vector that contains the prescribed values of H along the x boundary of the mesh,

$$\boldsymbol{B}_j \equiv \begin{bmatrix} l_{1,j}[(1-\Theta)H_{0,j} + \Theta H_{0,j+1}] \\ 0 \\ 0 \\ \vdots \\ 0 \\ u_{N,j}[(1-\Theta)H_{N+1,j} + \Theta H_{N+1,j+1}] \end{bmatrix},$$

and $\boldsymbol{M}_j$ is a tridiagonal $N \times N$ matrix

$$\boldsymbol{M}_j \equiv \begin{bmatrix} c_{1,j} & u_{1,j} & 0 & 0 & 0 & \cdots & 0 \\ l_{2,j} & c_{2,j} & u_{2,j} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & l_{N-1,j} & c_{N-1,j} & u_{N-1,j} \\ 0 & 0 & 0 & \cdots & 0 & l_{N,j} & c_{N,j} \end{bmatrix},$$

where

$$c_{i,j} = -\alpha \hat{v}_{i,j}, \tag{23a}$$

$$u_{i,j} = \tfrac{1}{2}\alpha(\hat{v}_{i,j} + \Delta_x \hat{b}_{i,j}), \tag{23b}$$

$$l_{i,j} = \tfrac{1}{2}\alpha(\hat{v}_{i,j} - \Delta_x \hat{b}_{i,j}). \tag{23c}$$

If the matrices in (22) can be determined, i.e. if the values of $\hat{r}_j$, $\hat{b}_{i,j}$, and $\hat{v}_{i,j}$ are known, the price of a contingent claim at time 0 can be obtained by iteratively solving the system of linear equations (22) backwards from the known time T payout vector $\boldsymbol{H}_{N+1}$. As the vector multiplying $\boldsymbol{H}_j$ is tridiagonal, the numerical solution to (22) can be coded very efficiently ($O(N)$); see e.g. Press *et al*. (1992: Chap. 2) for a discussion and specific algorithms. In Appendix A, we derive sufficient conditions for (22) to have a unique solution; these conditions will be satisfied by most realistic finite-difference meshes.

Contrary to a binomial tree where the value of a contingent claim on any given node can be determined from the state of only two nodes one time-step ahead (the 'up' and the 'down' nodes), the system of equations (22) generally links the value of $H_{i,j}$ to *all* interior values of H at time t_{j+1}, i.e. $H_{i,j} = F(H_{1,j+1}, \ldots, H_{N,j+1})$ for some function F. An exception occurs for the *explicit finite-difference scheme* ($\Theta = 1$) where the matrix multiplying $\boldsymbol{H}_j$ on the left-hand side of (22) is diagonal and $H_{i,j}$ consequently a function of only $H_{i+1,j+1}$, $H_{i,j+1}$, and $H_{i-1,j+1}$. According to (21), the equations for the explicit finite-difference scheme are

$$H_{i,j} = \frac{1}{1 + \hat{r}_j \Delta_t}\left[l_{i,j} H_{i-1,j+1} + H_{i,j+1}(1 + c_{i,j}) + u_{i,j} H_{i+1,j+1}\right]. \tag{24}$$

From (23a–c) we see that $l_{i,j} + (1 + c_{i,j}) + u_{i,j} = 1$, which allows for an interpretation of (24) as a trinomial tree with pseudo-probabilities of up, down, and center moves equal to $u_{i,j}$, $l_{i,j}$, and $1 - l_{i,j} - u_{i,j}$, respectively. While the explicit finite-difference grid has an attractive probabilistic interpretation and a simple causal structure, it unfortunately suffers from stability problems. In Appendix B, we derive conditions for the explicit finite-difference scheme to be stable; in most cases, these conditions are equivalent to all of the 'probabilities' $u_{i,j}$, $l_{i,j}$, or $1 + c_{i,j}$ being nonnegative. Due to the time-varying nature of $\hat{b}$ and $\hat{v}$, maintaining nonnegative probabilities at all nodes in the mesh puts heavy constraints on the spacing of the finite-difference mesh and, as in the binomial setting, turns out to interfere quite significantly with the fitting of the volatility smile. Consequently, the rest of this paper will *assume that* $\Theta \neq 1$ and instead focus on the Crank–Nicholson and implicit schemes which are known to have much better stability properties than the explicit scheme. Indeed, as a local harmonic analysis in Appendix B shows, both these schemes are unconditionally stable as long as $\hat{v}_{i,j} \geqslant 0$ for all i and j in the mesh. In most cases, we recommend the Crank–Nicholson scheme which has the best convergence properties of the three schemes. Specifically, the convergence order of the Crank–Nicholson scheme[7] is $O(\Delta_t^2)$, whereas both the explicit and the implicit schemes converge as $O(\Delta_t)$. All schemes converge as $O(\Delta_x^2)$ in x space.

3.2 Fitting of Bond Prices

As in the continuous-time case, we will assume the existence of a complete initial yield curve as given by prices of traded zero-coupon bonds maturing at all mesh times, $P_j = P(0, t_j)$, $j = 1, \dots, N+1$. As the strip of zero-coupon bonds can be interpreted as contingent claims with payout functions $g(S_{t_j}) = \$1$, their prices must satisfy the finite-difference equation (21). At time-step t_j consider the bond maturing one time-step ahead, $P(t_j, t_{j+1})$. Since, in our setting, bond prices are deterministic and thus independent of S (and x), (21) simplifies to

$$P(t_j, t_{j+1})\big(1 + \hat{r}_j \Delta_t\big) = P(t_{j+1}, t_{j+1}) = \$1, \qquad j = 1, \dots, N. \tag{25}$$

From (2) we know that

$$\frac{1}{P(t_j, t_{j+1})} = \frac{P(0, t_j)}{P(0, t_{j+1})} = \frac{P_j}{P_{j+1}},$$

so (with $P_0 = 1$)

$$\hat{r}_j = \frac{1}{\Delta_t}\left(\frac{P_j}{P_{j+1}} - 1\right), \qquad j = 0, \dots, N. \tag{26}$$

Not surprisingly, equation (26) is related to the continuous-time equation (3) through the finite-difference relation $\partial P(0, t_j)/\partial t \approx [P(0, t_{j+1}) - P(0, t_j)]/\Delta_t$.

3.3 Fitting of Asset Forwards

To match the drift of S, consider at time t_j a contract that pays out[8] $g(S_{t_{j+1}}) = S_{t_{j+1}}$ at the time-step t_{j+1} of the lattice. At node (x_i, t_j), the value of this contract is

$$H_{i,j} = S_i \exp\left(-\int_{t_j}^{t_{j+1}} \gamma(u)\, \mathrm{d}u\right) = \mathrm{e}^{x_i} \frac{\Gamma_{j+1}}{\Gamma_j}, \qquad i = 1, \dots, N, \quad j = 0, \dots, M, \tag{27}$$

where we have defined

$$\Gamma_j \equiv \exp\left(-\int_0^{t_j} \gamma(u)\, \mathrm{d}u\right).$$

Setting $H_{i,j} = S_i \Gamma_{j+1}/\Gamma_j$ and $H_{i,j+1} = S_i$ in the discretized PDE (21) and rearranging yields

$$\begin{aligned} &S_i\left[1 - \frac{\Gamma_{j+1}}{\Gamma_j}(1 + \hat{r}_j\Delta_t) - \alpha\hat{v}_{i,j}\left((1-\Theta)\frac{\Gamma_{j+1}}{\Gamma_j} + \Theta\right)\right] + (S_{i+1} + S_{i-1})\left[\tfrac{1}{2}\alpha\hat{v}_{i,j}\left((1-\Theta)\frac{\Gamma_{j+1}}{\Gamma_j} + \Theta\right)\right] \\ &+ (S_{i+1} - S_{i-1})\left[\tfrac{1}{2}\alpha\Delta_x\hat{b}_{i,j}\left((1-\Theta)\frac{\Gamma_{j+1}}{\Gamma_j} + \Theta\right)\right] = 0, \qquad i = 1, \dots, N, \quad j = 0, \dots, M. \end{aligned} \tag{28}$$

Now

$$S_{i+1} + S_{i-1} = S_i e^{\Delta_x} + S_i e^{-\Delta_x} = 2S_i \cosh \Delta_x, \qquad S_{i+1} - S_{i-1} = S_i e^{\Delta_x} - S_i e^{-\Delta_x} = 2S_i \sinh \Delta_x$$

so (28) becomes

$$\frac{1 - (\Gamma_{j+1}/\Gamma_j)(1 + \hat{r}_j \Delta_t)}{(1-\Theta)(\Gamma_{j+1}/\Gamma_j) + \Theta} + \alpha \hat{v}_{i,j}(\cosh \Delta_x - 1) + \alpha \hat{b}_{i,j} \Delta_x \sinh \Delta_x = 0. \tag{29}$$

Using the identity

$$\tanh \tfrac{1}{2}\Delta_x = \frac{\cosh \Delta_x - 1}{\sinh \Delta_x},$$

(29) can finally be rearranged as

$$\hat{b}_{i,j} = \frac{\Delta_x}{\Delta_t \sinh \Delta_x}\left(\frac{P_j/P_{j+1} - \Gamma_j/\Gamma_{j+1}}{(1-\Theta) + \Theta\Gamma_j/\Gamma_{j+1}}\right) - \frac{\hat{v}_{i,j}}{\Delta_x}\tanh\frac{\Delta_x}{2}, \qquad i = 1, \ldots, N, \quad j = 0, \ldots, M, \tag{30}$$

where we have used the result (26) to eliminate $\hat{r}_j$. It can easily be verified that $\hat{b}_{i,j} \to r_j - \gamma_j - \frac{1}{2}\hat{v}_{i,j}$ when Δ_x and Δ_t approach zero.

3.3 Fitting European Call Options

Equipped with (26) and (30), the discretization (21) will yield the correct forward stock and zero-coupon bond prices, irrespective of the volatility function $\hat{v}_{i,k}$. To determine the correct local volatilities, we assume the existence of observable call option prices with strikes and maturities spanning all nodes inside the upper and lower x boundaries of the finite-difference mesh (18a,b). Let $C_{\text{ini}}^{i,j}$ denote the time-0 observable value of a European call with strike $K = S_i = e^{x_i}$ and maturity of t_j, where $i = 1, \ldots, N$ and $j = 1, \ldots, M+1$.

While it would conceivably be possible to use brute-force trial-and-error techniques to back out a volatility function that correctly prices all calls $C_{\text{ini}}^{i,j}$ in the finite-difference mesh, this approach requires too much computational effort to be useful in practice. A significantly more efficient alternative is the so-called method of *forward induction* (Jamshidian 1991, Hull and White 1994), which avoids brute-force search by, in effect, introducing discrete-time versions of the Fokker–Planck forward equation (7) (or (14)). Rather than discretizing (7) or (14) directly, we will here use fundamental arguments to derive a discrete-time forward equation consistent with our backward discretization scheme (22). For this purpose, it is convenient and instructive to introduce the concept of *Arrow–Debreu securities*. To be specific, let $A_{\text{ini}}^{i,j}$ denote the time-0 price of a Arrow–Debreu security that at time t_j pays out \$1 if the asset price equals S_i and \$0 otherwise. To ensure correct pricing of bonds and stock forwards, the Arrow–Debreu securities must satisfy the following obvious constraints:

$$\sum_{i=0}^{N+1} A_{\text{ini}}^{i,j} = P_j, \qquad j = 0, \ldots, M+1, \tag{31a}$$

$$\sum_{i=0}^{N+1} A_{\text{ini}}^{i,j} S_i = S_{\text{ini}}\Gamma_j, \qquad j = 0, \ldots, M+1. \tag{31b}$$

It also follows from the definition of the European call payout function (8) that require that $D - U \geqslant 0$, or, after some manipulations,

$$\forall \omega: \quad 2\alpha \hat{v}_{k,j} + (1-2\Theta)\alpha^2 \left[\hat{v}_{k,j}^2 + \hat{b}_{k,j}^2 \Delta_x^2 + \cos \omega \Delta_x \left(\hat{b}_{k,j}^2 \Delta_x^2 - \hat{v}_{k,j}^2 \right) \right] \geqslant 0.$$

It follows that the finite-difference scheme (21) is stable if

$$0 \leqslant \Theta \leqslant \tfrac{1}{2} \left(1 + \frac{\frac{2}{\alpha} \hat{v}_{k,j}}{\hat{v}_{k,j}^2 + \hat{b}_{k,j}^2 \Delta_x^2 + \left| \hat{b}_{k,j}^2 \Delta_x^2 - \hat{v}_{k,j}^2 \right|} \right), \qquad k = 1, \ldots, N, \quad j = 0, \ldots, M. \tag{B.5}$$

From (B.5) we conclude that the finite difference scheme is *unconditionally stable (that is, stable for all* $\hat{v}_{i,j} \geqslant 0$*)* if $0 \leqslant \Theta \leqslant \frac{1}{2}$. Both the fully implicit ($\Theta = 0$) and the Crank–Nicholson ($\Theta = \frac{1}{2}$) finite difference schemes are thus unconditionally stable. For the explicit scheme ($\Theta = 1$), however, stability is only guaranteed if, for all $k = 1, \ldots, N$, $j = 0, \ldots, M$,

$$\frac{2}{\alpha} \hat{v}_{k,j} \geqslant \hat{v}_{k,j}^2 + \hat{b}_{k,j}^2 \Delta_x^2 + \left| \hat{b}_{k,j}^2 \Delta_x^2 - \hat{v}_{k,j}^2 \right|,$$

which is satisfied if

$$\hat{b}_{k,j}^2 \Delta_t \leqslant \hat{v}_{k,j} \leqslant \frac{\Delta_x^2}{\Delta_t}. \tag{B.6}$$

In practice, the main problem is the upper bound, which puts a constraint on the spacing of the time grid

$$\Delta_t \leqslant \frac{\Delta_x^2}{\hat{v}_{\max}},$$

where $\hat{v}_{\max}$ is the largest local volatility in the mesh. As the required time spacing is a quadratic function of the x-spacing, the number of time-steps necessary to ensure stability will frequently be impractically high.

ENDNOTES

[1] The standard notation for volatility, σ, is reserved for instantaneous volatility; see equation (1).

[2] As we shall see in Section 4, there are probably circumstances where *no* choice of K and T will give the correct value. Some practitioners appear to overcome this problem by using two constant volatilities for knock-out options: one volatility determines the probability of breaching the barrier, and one volatility is used to price the terminal call option. Needless to say, such an approach does not make any theoretical sense.

[3] The approach taken by Rubinstein (1994) does not suffer from this problem. However, Rubinstein's method ignores the intertemporal nature of the volatility smile and only fits the terminal time-slice of the tree to the volatility smile. As such, the method is not suited for options where the payout is a function of the path of the stock price process (as is the case for American

and most exotic options). Recently, Jackwerth (1996) and Brown and Toft (1996) have developed methods to extend the Rubinstein approach to multiple option maturities. Both approaches are quite different from the one taken in this paper, though.

[4] This equation was independently derived by Andreasen (1996).

[5] For example, one could set

$$x_{N+1} = x_{\text{ini}} + \int_0^{\mathrm{T}} [r(u) - \gamma(u)]\,\mathrm{d}u - \tfrac{1}{2}\sigma_x^2 T + Q_x \sigma_x \sqrt{T},$$
$$x_0 = x_{\text{ini}} + \int_0^{\mathrm{T}} [r(u) - \gamma(u)]\,\mathrm{d}u - \tfrac{1}{2}\sigma_x^2 T - Q_x \sigma_x \sqrt{T},$$

where Q_x is a positive constant, x_{ini} is defined in (19), and σ_x is some representative constant volatility. If σ_x is properly chosen, setting Q_x to a value of, say, 4 will assure that the (risk-neutral) probability of the terminal stock price S_T falling outside the borders of the mesh is less than 0.01%.

[6] This assumption is not critical but simplifies the exposition. If the original stock price lies between nodes, various interpolation techniques can be applied.

[7] The 2nd-order convergence of the Crank–Nicholson scheme in Δ_t is due to the fact that its estimator of the time-derivative is *central* (as it is evaluated in the middle of t_j and t_{j+1}).

[8] The value of this contract does *not* equal the stock price S_{t_j} as the holder of the contract is not entitled to any dividends paid over the interval $[t_j, t_{j+1}]$.

[9] As a side-effect of the finite size of the finite-difference mesh, these prescribed values will not exactly equal the observable market prices. As a consequence, (32b, c) will sometimes lead to small spikes in the prices of Arrow–Debreu securities that pay out in the extreme upper and lower x boundaries of the mesh. This is no cause of concern as the absolute magnitude of these prices are typically $\ll 10^{-6}$ and hence have practically no influence on realistic option prices. At the price of a slight loss of accuracy, the spikes can be removed by abandoning (32b, c) and using the observable prices instead.

[10] The matrix Φ_j determines the scale of the quadratic program and only affects volatilities if the constraints are binding. In most cases, it is sufficient to set Φ_j equal to the identity matrix. For improved accuracy, particularly for long-term deals ($T > 5$ years), it is sometimes better to pick the call option payout matrix, $\{\varphi_{r,c}\} = \max(S_c - S_{r-1}, 0) = \max(\mathrm{e}^{x_c} - \mathrm{e}^{x_{r-1}}, 0),\ r, c = 1, \ldots, N$. With this transformation, the quadratic program will optimize on option prices rather than Arrow–Debreu prices.

[11] Alternatively, one could interpolate directly on call option prices. Unless one is quite careful, this will tend to result in irregular implied volatilities, particularly for high and low strikes (where the sensitivity of option prices to implied volatility is low). Moreover, as noted in Andreasen (1996), it is easier to extrapolate and smooth on surfaces that are relatively flat (which presumably is the case for implied volatilities). Notice, however, that if stock forwards are themselves not smooth, as would for example be the case if dividends are modeled as discrete-time lump sum payments, it would not be appropriate to model implied volatilities (or call prices, for that matter) as smooth functions of T and K. Such cases can be handled by normalizing strikes with time-0 forward values

calibration of short-rate interest rate models to market data which is normally performed either by binomial trees (Jamshidian 1991) or by modified explicit finite-difference schemes (Hull and White 1994).

REFERENCES

Andreasen, J., Implied modelling: stable implementation, hedging, and duality, Working Paper, University of Aarhus, 1996.

Arnold, L., *Stochastic Differential Equations: Theory and Practice*, John Wiley & Sons, 1974.

Avallaneda M., C. Friedman, R. Holmes, and D. Samperi, Calibrating volatility surfaces via relative-entropy minimization, Working Paper, Courant Institute, New York University, 1996.

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

Black, F. and M. Scholes, The pricing of options and other corporate liabilities, *Journal of Political Economy*, **81** (May–June 1973), 637–659.

Beckers, S., The constant elasticity of variance model and its implications for options pricing, *Journal of Finance*, **35** (1980), 661–673.

Boyle, P. and S. Lau, Bumping up against the barrier with the binomial method, *Journal of Derivatives*, **4** (1994), 6–14.

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

Brennan, M. and E. Schwartz, Finite difference methods and jump processes arising in the pricing of contingent claims: a synthesis, *Journal of Financial and Quantitative Analysis*, **13** (September 1978), 462–474.

Brown, G. and K. Toft, Constructing binomial trees from multiple implied probability distributions, Working Paper, University of Texas at Austin, 1996

Chriss, N., Transatlantic trees, *Risk* (July 1996), 45–48.

Courtadon, G., A more accurate finite difference approximation for the valuation of options, *Journal of Financial and Quantitative Analysis*, **17** (1982), 697–705.

Cox, D. and H. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, 1965.

Cox, J. and S. Ross, The valuation of options for alternative stochastic processes, *Journal of Financial Economics*, **3** (March 1976), 145–166.

Derman, E. and I. Kani, Riding on a smile, *Risk* (February 1994), 32–39.

Derman, E., I. Kani, and N. Chriss, Implied trinomial trees of the volatility smile, Working Paper, Goldman Sachs, 1996.

Dewynne, J. and P. Wilmott, Partial to the exotic, *Risk* (March 1993), 38–46.

Dewynne, J., S. Howison, and P. Wilmott, *Option Pricing*, Oxford Financial Press, 1993.

Dierckx, P., *Curve and Surface Fitting with Splines*, Oxford Science Publications, 1995.

Dupire, B., Pricing with a smile, *Risk* (January 1994), 18–20.

Ecker, J. and M. Kupferschmid, *Introduction to Operations Research*, John Wiley & Sons, 1988.

Geske, R. and K. Shastri, Valuation of approximation: a comparison of alternative approaches, *Journal of Financial and Quantitative Analysis*, **20** (March 1985), 45–72.

Hull, J. and A. White, The pricing of options on assets with stochastic volatilities, *Journal of Finance*, **3** (1987), 281–300.

Hull, J. and A. White, Valuing derivative securities using the explicit finite difference method, *Journal of Financial and Quantitative Analysis*, **25** (March 1990), 87–100

Hull, J and A. White, Numerical procedures for implementing term structure models I: Single-factor models, *Journal of Derivatives*, **1** (Fall 1994), 7–16.

Jackwerth, J., Generalized binomial trees, Working Paper, University of California at Berkeley, 1996.

Jackwerth, J. and M. Rubinstein, Recovering probability distributions from option prices, *Journal of Finance*, **51** (December 1996), 1611–1631.

Jamshidian, F., Forward induction and construction of yield curve diffusion models, *Journal of Fixed Income*, **1** (June 1991), 62–74.

Karatzas, I. and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer Verlag, 1991.

Lagnado, R. and S. Osher, Reconciling differences, *Risk* (April 1997), 79–83.

Merton, R., Theory of rational option pricing, *Bell Journal of Economics and Management Science*, **4** (Spring 1973), 141–183.

Merton, R., Option pricing when underlying stock returns are discontinuous, *Journal of Financial Economics*, **3** (1976), 124–144.

Platen, E. and M. Schweizer, On smile and skewness, Working Paper, School of Mathematical Sciences, Australian National University, 1994.

Press, W., S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992.

Rubinstein, M., Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 13, 1976, through August 31, 1978, *Journal of Finance*, **40** (June 1985), 455–480.

Rubinstein, M., Implied binomial trees, *Journal of Finance*, **49** (July 1994), 771–818.

Rubinstein, M., As simple as 1-2-3, *Risk* (January 1995), 44–47.

Rubinstein, M. and E. Reiner, Breaking down the barriers, *Risk* (September 1991), 28–35.

Shimko, D., Bounds of probability, *Risk* (April 1993), 33–37.

Zvan, R., P. Forsyth, and K. Vetzal, PDE methods for pricing barrier options, Working Paper, University of Waterloo, 1997

APPENDICES

Appendix A. Sufficient conditions for invertibility of (22)

In (22), consider the tridiagonal matrix

$$(1 + \hat{r}_j \Delta_t)\boldsymbol{I} - (1 - \Theta)\boldsymbol{M}_j.$$

It is well-known that a sufficient condition for a tridiagonal matrix to be invertible is that it is *diagonally dominant*, i.e. the absolute value of the diagonal element in each