Valuing American options in the presence of user-defined smiles and time-dependent volatility: scenario analysis, model stress and lower-bound pricing applications.

Peter Jäckel
Riccardo Rebonato
Quantitative Research Centre, The Royal Bank of Scotland
135 Bishopsgate, London EC2M 3UR

August 7th, 2000

Abstract

A method is presented for the valuation of American style options as a function of an exogeneosly assigned future implied volatility surface. Subject to this specification, this is done independently of any assumptions about a stochastic process of the underlying asset. The scope and applicability of the method are analyzed in detail. The approach should have applications in risk management, and to estimate a lower bound price for the American option.

1 Introduction and motivation

The problem of valuing American FX and equity options has received a lot of attention and a variety of methods, of different generality and computational effectiveness, have been proposed. If some very restrictive assumptions on the process of the underlying are made, the procedure of Derman and Kani [DKZ96] can even provide (or can be extended to provide) a valuation in the case of a "smiley" volatility surface by assuming a deterministic instantaneous volatility function depending on the current spot value and calendar time. For financially non-trivial applications (time- and/or state-dependent volatility, time-dependent interst rates etc.) modifications of the original finite-differences inspired approach pioneered by Brennan and Schwartz [BS77], Parkinson [Par77], and others, is still used because of its versatility. See Ingersoll [Ing98] for one such approach and for a review of the current methods. What, to the best knowledge of the authors, is not available yet is a method which is capable of solving the accompanying free-boundary problem when the user requires the dynamics of the underlying to reflect not only today's smile surface, but also an exogenously specified evolution of the surface itself. This second point is absolutely crucial: the Derman and Kani approach and related methods[1] allow the unique distillation of the local volatility surface compatible with the observed market prices of options when the underlying dynamics are described by a process of the form

$$dS = (r - d)Sdt + \sigma(S,t)\,dW\ . \tag{1}$$

In the equation above, S is the asset price, r the short rate, d the dividend yield, σ the volatility, and dW the increment of a standard Wiener process. Once the assumption embodied by equation (1) is made, however, the future shape of the smile surface is uniquely determined[2]. Notice that, if one looks at the problem in this light, the implied volatility can no longer be regarded as some suitable average of the (square of) the future locality volatility, as it is the case for a purely time-dependent volatility. Rather, it is solely the input to feed into a market-agreed algorithm (the Black-Scholes formula) to produce prices. In the presence of smiles, it is more to be considered as "the wrong number to put in the wrong formula to obtain the right price" [Reb99] than as the constant volatility of an underlying geometric Brownian motion.

A fundamental problem for a local volatility approach such as the one described by equation (1) is that the typical shapes for the equity and FX volatility surfaces, which have been observed for at least a decade, are often predicted by a local-volatility solution to change radically or even to virtually disappear (see [Reb99] for a detailed discussion of the future implied volatility in a Derman-Kani world for several shapes of the initial implied volatility surface both in the FX and in the equity context). In general, one can say that the implied volatility surface tends to evolve in a strongly time-inhomogeneous way when the underlying is driven by a process such as in equation (1). Since the price of an option is equal to the set-up cost of the hedging portfolio, this failure to reproduce plausible *future* implied volatility surfaces has obvious and strong repercussions on the option price today.

Other, more general, processes, such as, for instance, jump-diffusions can easily produce a much more time-homogeneous evolution for the implied volatility surface. Unfortunately, they do

[1] From the computational point of view, the Derman and Kani approach may not be the most efficient method. The solution, however, is unique (as long as the drift of the process is specified and a continuum of prices across strikes and maturities is available). Any other method, such as the one proposed by Rebonato [Reb99] capable of providing *a* solution of better numerical quality is therefore guaranteed to provide *the* solution.

[2] This future implied volatility surface is conceptually best obtained by placing oneself at a future point in time in the local volatility space $(S(t), t)$, calculating future prices of plain-vanilla options from this point and implying the corresponding Black-Scholes volatility.

not readily lend themselves to the finite difference or tree-based numerical techniques typically used for the evaluation of American or compound options, and for which the approach subsumed by equation (1) is ideally suited.

In reality, given a specific market, the user would like to have a model capable not only of recovering today's option prices, but also to reflect his view about the "sticky" or "floating" nature of the smile. (The term "sticky delta" is sometimes used [Rei98] instead of "floating"). In other terms, the user would like to be able to assess the quality of a model on the basis of its ability to reflect available econometric data and/or the trader's intuition about how the smile will evolve as the underlying moves over time. It must be stressed that this information is not unambiguously recoverable from today's observed prices, and an element of financial judgment is always required in addition to the current price information to pin down the dynamics of the process.

If, however, this is the case, a useful new insight is available: if the quality of a model is determined by its ability to reproduce future smile surfaces (i.e. future prices of plain-vanilla options for different strikes) congruent with the user's expectation, shouldn't one concentrate directly on these plain-vanilla prices, rather than proceeding via the intermediate model-building stage? This could be of use if one could express prices of non-plain-vanilla options as linear combinations of calls and puts. How this can be done for known-boundary problems (such as double knock-out barriers or up-and-out call options with optional rebate etc.) has been shown in detail by Rebonato [Reb99] for sticky and floating smiles. For this class of problems, the solution is asymptotically exact. The purpose of this paper is to provide an efficient computational approach to extend the same method to free-boundary problems, such as the ones encountered in the valuation of American options. As shown in detail in the following, the attempt to find a numerical approximation to the free-boundary problem in question by using a linear combination of plain-vanilla options is equivalent to choosing the Black-Scholes solutions to the European call and put option problem as the basis function set for the discretisation of the dynamical evolution of the option value. Despite the fact that, for a general non-diffusive process for the underlying, these basis functions do not provide a solution for the problem at hand, they presumably provide a good approximation to the unknown solution. Therefore, as we shall see, even when only a few plain-vanilla options are included, the approximation already gives a price estimate of excellent accuracy in those situations where comparison with other pricing methods is available.

Before delving into the details of the method, it is important to stress the scope of the method:

- no assumption is made that the Black-Scholes formula should be the discounted expectation of the pay-offs produced by a particular process: in agreement with the market, no consensus is necessary about the underlying process in order to adopt the Black-Scholes formula beyond its range of applicability (as it is routinely done whenever a "smiley" volatility is quoted). In this sense the celebrated formula simply becomes a black (Black) box capable of producing prices once fed a single input (the implied volatility);

- the user is required, in the present approach, to have an opinion about the future implied volatility surface. This is, however, unavoidable in all approaches. The only difference is whether the user prefers an opaque model to make this choice on his behalf, or whether he prefers to impose it directly, on the basis of historical analysis or of his market knowledge;
- the approach, for reasons discussed in section 2, can only provide a lower bound for the price a risk-averse trader should quote for an American option. Alternatively, it can offer a useful tool for risk management purposes, in order to evaluate quickly and efficiently portfolios of American options under user-assigned future smile surfaces.
- once the choice about the nature of the smile has been made (*e.g.* sticky or floating), closed form expressions are available for the delta, gamma, and vega sensitivities, just as they are for the Black-Scholes pricing formula.

The second point might appear counter-intuitive: the user seems to be required to know what future (plain-vanilla) prices will be in order to be able to determine a spot (non-plain-vanilla) price. In reality, the user is really requested to make a price in the future *shape* (as opposed to the future *level*) of the volatility surface: rigid moves up and down of the volatility curves are relatively easy to hedge against by means of the traditional vega hedging, and are therefore unlikely to be reflected in the market prices of plain-vanilla options. What the user really needs to control is that the future wings of the volatility surface will resemble what has historically been observed in the past for similar degrees of out-of-the-moneiness[3]. This state of affairs has an interesting and direct counterpart in the interest-rate area: in the case of Libor products an infinity of specifications for the instantaneous volatility functions of the forward rates can produce the same fit to today's prices of discount bonds and caplets, but they all imply different future term structure of volatilities (see Rebonato [Reb98] and [Reb99]). Once again, the user has to make a choice between different models on the basis of plausibility of these *future* prices (of caplets). Also in this case, the quality of an interest-rate model can be assessed on the basis of its ability to reproduce a future market observable (the term structure of volatilities).

The remainder of this paper is organised as follows. In section 2 we shall carefully discuss the scope and the range of applicability of the model. Next, in section 3, we will describe the method in detail. Then, in section 4, comparison will be made with those cases (constant and time-dependent volatility) where finite differences methods can provide accurate numerical solutions.

[3]Needless to say, if the trader had different views about the future evolution of the smile surface, these views, rather than the historical past, should be reflected in the future smile.

Beyond this, the effect of different assumed future smile dynamics will be discussed and we will comment on the accuracy of the delta, gamma, and vega statistics. Following that, the behaviour of American-style options near the exercise boundary will be analysed in section 5. Finally, the conclusions will be reported in section 6.

2 Applicability

The method presented in this paper assumes that the user can freely assign a *future* implied volatility surface. Since this is an unconventional procedure, we must explore carefully to what extent, and in what contexts, this is a legitimate exercise. Since we shall show that the approach can have useful but different applications for risk managers and traders, we shall refer collectively to 'users' in the first part of the section, before the domain of applicability of the technique has been made clear. In order to address this question we begin by defining *admissible* smile surfaces. These are defined to be surfaces such that, at any point in time and for any strike and maturity, the following relationships hold true for the value C of a European call option:

$$\frac{\partial^2 C}{\partial K^2} > 0 \tag{2}$$

$$\frac{\partial C}{\partial K} < 0 \tag{3}$$

$$\frac{\partial C}{\partial T} > 0 \tag{4}$$

In the expressions above, K is the strike and T is the option maturity. These conditions are model- (process-) independent and, if violated, would expose the trader to arbitrage whatever the process for the underlying. They are in the same spirit[4] as Merton's conditions in his theory of rational option pricing [Mer73]. With obvious notation, the buy-and-hold, model-independent arbitrage strategies that would ensure risk-less profits if any of the conditions were violated are:

Buy $C(K + \Delta K)$, Buy $C(K - \Delta K)$, Sell $2 \cdot C(K)$ [equation (2)]

Buy $C(K)$, Sell $C(K + \Delta K)$ [equation (3)]

Buy $C(T + \Delta T)$, Sell $C(T)$ [equation (4)]

[4] Strictly speaking, equation (4) does not have to hold for arbitrary interest and dividend rates. In the simplified case of zero interest rates and no dividends, however, it is the easiest way to express the principle that *our knowledge about the distant future can never be more than our knowledge about the immediate future* in a single formula.

In the following we shall assume that the user of the method will always assign admissible future implied volatility surfaces. Another condition that must hold true whatever the 'true' process of the underlying process is the link between the spot and the forward prices:

$$F(t,T) = S(t)e^{(r-d)(T-t)} \tag{5}$$

where $F(t,T)$ indicates the time-t value of the T-expiry forward price, $S(t)$ denotes the stock price at time t, and r and d are the interest rate and dividend yield, respectively[5]. Also in this case, a simple buy-and-hold, process-independent strategy (the cash-and-carry arbitrage) punishes traders who (in a perfect, friction-less market) enter forward transactions at a different price. Note, however, that, as long as the prices of options are obtained from an admissible smile surface (via application of the Black-and-Scholes formula), the forward condition is automatically guaranteed to be satisfied. Therefore, because of call/put parity, an admissible smile surface automatically also ensures that the forward contract is correctly priced, and the user specifying an admissible future implied volatility surface does not have to take into account the forward condition separately.

Let us now consider a two-period trading horizon and one such admissible smile surface today, i.e. today's prices of calls and puts of all strikes maturing at times t_1 and t_2 are assumed to be available. From the call prices, $\{C\}$, and using equation (2), one can obtain the unconditional (risk-neutral[6]) probability densities $p\left(S_{t_0} \to S_{t_i}\right), i = 1, 2$:

$$\frac{\partial^2 C(S{=}S_{t_0}, K, t, t_i)}{\partial K^2} = p\left(S_{t_0} \to S_{t_i}{=}K\right) \tag{6}$$

In the expression above, $p\left(S_{t_0} \to S_{t_i}{=}K\right)$ indicates the probability for the stock price to move from the value S_{t_0} today to value K at time t_i. To the extent that prices of calls of all strikes and maturities are given by the market, the unconditional probability density can therefore be assumed to be market-given: the user has no freedom to modify it, or to have 'views' about it.

For future discussion, it is essential to make a distinction at this point between the 'true' price functional, and the Black-and-Scholes pricing formula. The former depends on today's value of the stock price, the residual time to maturity, the strike of the option, an unknown set of parameters describing the 'true' dynamics (diffusion coefficients, jump amplitudes, etc.), and, possibly, on the past history. The parameters describing the process of the underlying (volatility, jump frequency, jump amplitude, etc.) can, in turn, be themselves stochastic. However, they are all, obviously, strike-independent. The unknown 'true' parameters and the full history at time t will be symbolically denoted by $\{\alpha_t\}$ and $\{\boldsymbol{F}_t\}$, respectively. More technically, $\{\boldsymbol{F}_t\}$ is the filtration generated by the evolution of the stock price and of whatever stochastic parameters describe its process.

[5]Interest rates and dividend yields need not be constant. The generalisation is trivial. Also, the same expression applies to forward FX rates, with the domestic and foreign rates replacing r and d, respectively.

[6]All the probability densities referred to in this section are risk-neutral.

The Black-and-Scholes formula, on the other hand, depends on today's value of the stock price, the residual time to maturity, the strike of the option and a single strike-dependent parameter (the implied volatility). If, with obvious notation, the true functional is denoted by $C(S_t, T - t, K, \{\alpha_t\}, \{\boldsymbol{F}_t\})$ and the Black-and-Scholes formula by $BS(S, T - t, K, \hat{\sigma}(t, T, K))$, then we would like to be able to write

$$C(S_t, T-t, K, \{\alpha_t\}, \{\boldsymbol{F}_t\}) = BS(S, T-t, K, \hat{\sigma}(t, T, K)) \tag{7}$$

As of today this expression is certainly well defined, since, from the knowledge of the evolution up to today of the stock process and of its stochastic parameters (if any), we know what values to input in the true pricing functional. As a consequence, as of today, even if the user does not know the true functional C, he can still write

$$\frac{\partial^2 C(S_0, K, T, \{\alpha_t\}, \{\boldsymbol{F}_t\})}{\partial K^2} = \frac{\partial}{\partial K}\left[-N(h_2) + \frac{\partial BS}{\partial \hat{\sigma}}\frac{\partial \hat{\sigma}}{\partial K}\right] = p\left(S_0 \to S_T{=}K\right) \tag{8}$$

where use has been made of the fact that $\frac{\partial BS}{\partial K} = -N(h_2)$, and h_2 is given by

$$h_2 = \frac{\ln(S/K) + (r-d)T - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\ . \tag{9}$$

It is important to stress again the fact the fundamental difference between Equation (8) and equation (6): if the trader assumes to know how the function $\hat{\sigma}(0, T, K)$ changes with strike then the risk-neutral unconditional probability densities for all maturities can be obtained analytically *even if the true functional C is not known*.

What one would like to be able to do is to repeat the same reasoning forward in time, and obtain, by so doing, expressions for the *conditional* probability densities. In other words, it would be tempting to reason as follows: using Kolmogorov's equation, one might like to be able to write that these unconditional probability densities at times t_1 and t_2 are linked by

$$p\left(S_{t_0} \to S^*_{t_2}\right) = \int p\left(S_{t_0} \to S'_{t_1}\right) p\left(S_{t_1} \to S^*_{t_2} \middle| S_{t_1}{=}S'_{t_1}\right) dS'_{t_1} \tag{10}$$

where the quantity $p\left(S_{t_1} \to S^*_{t_2} \middle| S_{t_1}{=}S'_{t_1}\right)$ now denotes the conditional probability density of the stock price reaching S^* at time t_2, given that it had value S'_{t_1} at time t_1. Notice, however, that the integral (10) can be used to obtain information about future conditional price densities if and only if a future state of the world is fully determined by the realisation of the stock price. In general, however, this will not be the case. If, for instance, the 'true' process for the stock price were driven by a stochastic volatility, expression (10) would require a double integration over future price and volatility values:

$$\begin{aligned} p\left(S_{t_0} \to S^*_{t_2}, \sigma_{t_0} \to \sigma^*_{t_2}\right) \;=\; & \iint p\left(S_{t_0} \to S'_{t_1}, \sigma_{t_0} \to \sigma'_{t_1}\right) \cdot \\ & p\left(S_{t_1} \to S^*_{t_2}, \sigma_{t_1} \to \sigma^*_{t_2} \middle| S_{t_1}{=}S'_{t_1} \wedge \sigma_{t_1}{=}\sigma'_{t_1}\right)\, dS'_{t_1} d\sigma'_{t_1} \end{aligned} \tag{11}$$

where $p\left(S_{t_0} \rightarrow S'_{t_1}, \sigma_{t_0} \rightarrow \sigma'_{t_1}\right)$ is now the unconditional probability density that the stock price should have value S'_{t_1} at t_1 and its volatility has become σ'_{t_1}, and

$$p\left(S_{t_1} \rightarrow S^*_{t_2}, \sigma_{t_1} \rightarrow \sigma^*_{t_2} \middle| S_{t_1}{=}S'_{t_1} \wedge \sigma_{t_1}{=}\sigma'_{t_1}\right)$$

is the conditional probability of reaching state $S^*_{t_2}$ and $\sigma^*_{t_2}$ at time t_2, given that the stock price and the volatility had values S'_{t_1} and σ'_{t_1} at time t_1.

Since such integral expressions quickly become more cumbersome if one wants to include explicitly all the stochastic variables that specify a future state, it is more convenient to speak simply of conditional and unconditional probability densities of reaching states. Also, if, instead of a continuum of states, one assumes that there exists only a discrete set of possible future states, one can re-state more concisely Equation (11) in matrix form. The equivalent, discrete-price, expression in terms of transition matrices is given by

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1n} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{n1} & \pi_{n2} & \dots & \pi_{nn} \end{pmatrix} \cdot \begin{pmatrix} p^1_1 \\ p^1_2 \\ \vdots \\ p^1_n \end{pmatrix} = \begin{pmatrix} p^2_1 \\ p^2_2 \\ \vdots \\ p^2_n \end{pmatrix} \tag{12}$$

or, more concisely,

$$\mathbf{\Pi} \cdot \boldsymbol{p}^1 = \boldsymbol{p}^2 \tag{13}$$

In equation (12), π_{jk} is the conditional probability of going from state k at time t_1 to state j at time t_2. These conditional probabilities are not all independent. Since, from a given state, the total probability of reaching some other state is unity, one must impose a probability normalisation condition from each parent state. This provides n equations ($\sum_k \pi_{jk} = 1$). It is therefore easy to see that equations (12) together with the normalisation condition provide $2n$ constraints, which do not uniquely specify the n^2 elements of the matrix $\mathbf{\Pi}$. As a consequence, a number of transition matrices satisfy equation (12) — which is another way of saying that a variety of processes can account for today's option prices (as embodied by the vectors $\boldsymbol{p}^1$ and $\boldsymbol{p}^1$) by the discretised equivalent of (6). Similarly, an infinity of conditional probability densities satisfy the Kolmogorov equation (11), with the link between the probability density and the prices explicitly given by

$$p\left(S_{t_1} \rightarrow S^*_{t_2} \middle| S_{t_1}{=}S'_{t_1}\right) = \frac{\partial^2}{\partial K^2} C\left(S{=}S'_{t_1}, K{=}S^*_{t_2}, T{=}t_2{-}t_1\right) \ . \tag{14}$$

As long as Equations (11) or (12) are satisfied, an infinity of future smile surfaces are therefore compatible with today's prices of plain-vanilla options. If the admissible future smile surface assumed by a trader (Trader 1) belongs to this set, then he can rest assured that no static, model-independent, strategy can arbitrage his prices. Another trader (Trader 2), with superior and correct knowledge about the 'truth' of a specific process, can, of course arbitrage Trader 1, but he can only do that by engaging in a dynamic, model-dependent strategy. Since we consider, in this section, the situation of users who have views about future smile surfaces, but are 'agnostic' about true models, the possibility of model-driven arbitrage is not our main concern. It is easy to see, however, that a trader can believe that a single (admissible) future volatility surface will prevail at time t if and only if he believes that all the (possibly unknown) parameters driving the process for the stock price (volatility, jump amplitude ratio, jump frequency, etc.) fall in either of two classes :

i) either they are fully deterministic (i.e. either constant, or, at most, purely dependent on time);

ii) or their value at time t depends at most on the time-t realisation of the stock price itself.

Since the most common processes are jump-diffusions, let us analyse in this light the implications of i) and ii) above. If the process is a pure diffusion, then the volatility must have the form σ_0 or $\sigma(t)$, for case i) to apply, and $\sigma(S_t, t)$ for case ii). If the volatility is of the form σ_0 or $\sigma(t)$, then today's prices fully determine the future volatility surface (which, incidentally, can display no smiles). If the volatility is of the form $\sigma(S_t, t)$ we are back to the Derman-Kani-Dupire-Rubinstein (DKDR) solution, which, as it is well known, is unique. Once again, the future implied volatility surface is therefore uniquely determined by today's prices. If the user assigns a single future implied volatility surface different from the DKDR one, he is implicitly assuming that the process *cannot* be a diffusion without jumps: a stochastic volatility is not compatible with a unique future smile surface; a volatility of the form σ_0 or $\sigma(t)$ allows no smiles whatsoever; and a volatility of the form $\sigma(S_t, t)$ is only compatible with the DKDR implied smile surface. If the user insists that a single future implied volatility surface will prevail, and that the process is no more complex than a jump-diffusion, then he must believe that, on top of a diffusion of type i) or ii) there must exist a jump component with parameters $\{\boldsymbol{j}\}$ either constant $\{\boldsymbol{j}_0\}$, or time dependent, $\{\boldsymbol{j}(t)\}$, or, at most, dependent on time and on the realisation of the stock price $\{\boldsymbol{j}(S_t, t)\}$. These semi-agnostic views are logically consistent with assigning a single future implied volatility surface. Unfortunately, and for totally un-related technical reasons, in the presence of jumps[7] the replication method presented below becomes inapplicable.

Despite these considerations, there are still important useful applications of the method presented in the following sections. To begin with the 'user' might be a risk manager, who is not interested in arbitrage-free option pricing, but in the impact of different possible implied volatility scenarios on a portfolio of trades. The only constraint the risk manager would have to satisfy would

[7] The technique presented in the following assumes, at least conceptually, liquidation of a portfolio of plain-vanilla options 'as soon as' a certain barrier is crossed. Even with continuous trading, the likelihood of exceeding this barrier cannot be made arbitrarily small if finite jumps are possible.

be the admissibility of the implied volatility smile surface. For this purpose the method presented in the following provides an efficient and powerful tool for scenario and model stress analysis.

Alternatively, the method presented below could be used for trading purposes: a trader who wished to use the approach would have to abandon the idea of specifying a *single* future smile surface, and will have to provide, instead, a variety of possible future smile surfaces constrained to be admissible and to belong to the set of surfaces obtainable from the Kolmogorov equations (11). In the absence of a liquid market in forward-starting options (i.e. given the incompleteness of the volatility market), these joint conditions would be sufficient to ensure that no *model-independent* dynamic trading strategy can arbitrage any of the prices implied by the future volatility surface. A separate note shows how these Kolmogorov-compatible future implied volatility surfaces can be obtained in an efficient way. Due to risk aversion, and, consequently, to the existence of market price(s) of risk, the average of the prices obtained using the different scenarios is clearly not equal to the 'correct' price; it can, however, constitute a lower or upper bound for the price for the risk averse option buyer or seller, respectively. The dispersion in prices would also assist the trader in making a price, taking his risk aversion into account.

3 The method in general

3.1 Replication of American options

The algorithmic description of static replication of derivative contracts with known boundary conditions such as continuous double barrier options, single barrier contracts with rebates, etc., is well known [Reb99] : the key is to use several plain-vanilla options maturing at different times to approximate the desired profile on the boundary as closely as possible. A limiting process then produces an asymptotic solution for the continuous problem. For discretely monitored contracts, the static replication method requires that not only the value of a portfolio at a given point in spot-time space be adjusted by including additional suitable plain-vanilla options, but also that the slope of the portfolio value with respect to the underlying asset value be matched. The idea of matching the slope of the portfolio profile, i.e. the delta of the replication, leads to the replication of free boundary products where the boundary level itself is unknown, but boundary conditions are known for the derivative with respect to the underlying asset.

The valuation of American-style options is complicated by the fact that there is no *a priori* knowledge of the location of the boundary. Rather, we only know that the value of the contract is never to be less than its intrinsic value, and that the derivative with respect to the underlying is to be continuous. For an American-style call option[8], this tells us that the boundary is where

[8] *Mutatis mutandis*, the same reasoning clearly applies to put options.

the delta first becomes equal to unity. The key to the replication of American options is thus to set up a portfolio of plain-vanilla options with maturities on a selected set of n times $\{t_i\}$ for $i = 1 \ldots n$ such that at each time slice t_i the portfolio is never worth less than the intrinsic value of the contract, and, at the point where it becomes equal to the intrinsic value, it has unit slope on the outside of the boundary.

The starting point is then a given discretisation of time between inception and maturity of the option. The last point in this discretisation is to be chosen as maturity, i.e. $t_n = T$. In order to replicate an American call option struck at K and maturing at T and to calculate its value $C_{\text{Am}}(S_{t_0}, t_0, K, T)$ at time t_0 when the underlying asset has value S_0, we begin to set up our replication portfolio by including a European-style call option also struck at K and maturing at T. In our notation, the value of this European call option at any time t for a given value of the underlying S_t is denoted by $C_{\text{Eur}}(S_t, t, K, T)$. At time $t_n = T$, the value of the American option, of the European option, and of the intrinsic value all coincide:

$$C_{\text{Am}}(S_{t_n}, t_n, K, t_n) = C_{\text{Eur}}(S_{t_n}, t_n, K, t_n) = \max\left(S_{t_n} - K, 0\right) \tag{15}$$

One can then move backwards by one time step to time t_{n-1}. At this time one can write for the value of the doubly exercisable option

$$\hat{C}_{\text{Am}}(S_{t_{n-1}}, t_{n-1}, K, t_n; n) = \max\left(\breve{\mathrm{E}}_{t_{n-1}}\left[\max\left((S_{t_n} - K), 0\right)\right], \max\left(S_{t_{n-1}} - K, 0\right)\right) \ . \tag{16}$$

In equation (16) $\breve{\mathrm{E}}_t\left[\cdot\right]$ denotes the discounted expectation as seen from time t and, in our notation, $\hat{C}_{\text{Am}}(S_{t_{n-1}}, t_{n-1}, K, t_n; n)$ is the approximate value for the American call option as given by the replication procedure using n time steps. Clearly, the ultimate aim is the asymptotic convergence of the approximate value of the option at inception, i.e.

$$\lim_{n\to\infty} \hat{C}_{\text{Am}}(S_0, t_0, K, T; n) = C_{\text{Am}}(S_0, t_0, K, T) \ . \tag{17}$$

The quantity $\breve{\mathrm{E}}_{t_{n-1}}\left[\max\left((S_{t_n} - K), 0\right)\right]$ is by definition given by $C_{\text{Eur}}(S_{t_{n-1}}, t_{n-1}, K, t_n)$, i.e. it is the output of the Black-Scholes formula with the appropriate future implied volatility prevailing at time t_{n-1} for time to maturity $\Delta T_n = (t_n - t_{n-1})$. It must be stressed that using the Black-Scholes formula in conjunction with a smiley implied volatility does not imply endorsement of any of the assumptions underpinning the Black-Scholes world. If the user believed, for instance, volatility smiles to be, say, homogeneous in time and floating in spot, he would apply today's one-period implied volatility as read from the appropriate level of at-the-moneiness. As mentioned above, in the presence of smiles this information is not contained in an unambiguous manner even in all of today's spot prices corresponding to the whole strike/maturity surface.

If the yield one would earn when holding the underlying asset is positive, i.e. for positive dividend yield on equity or foreign interest rates for FX options, there will be a spot level $S^*_{t_{n-1}}$

beyond which the portfolio will be worth less than the intrinsic value of the American option that is to be replicated. This intersection of the value of the European option and the intrinsic value on time slice t_{n-1} can be located very efficiently, for instance, by the aid of the Newton iteration method. Once we have knowledge of the precise value of $S^*_{t_{n-1}}$ and of the delta of the portfolio as

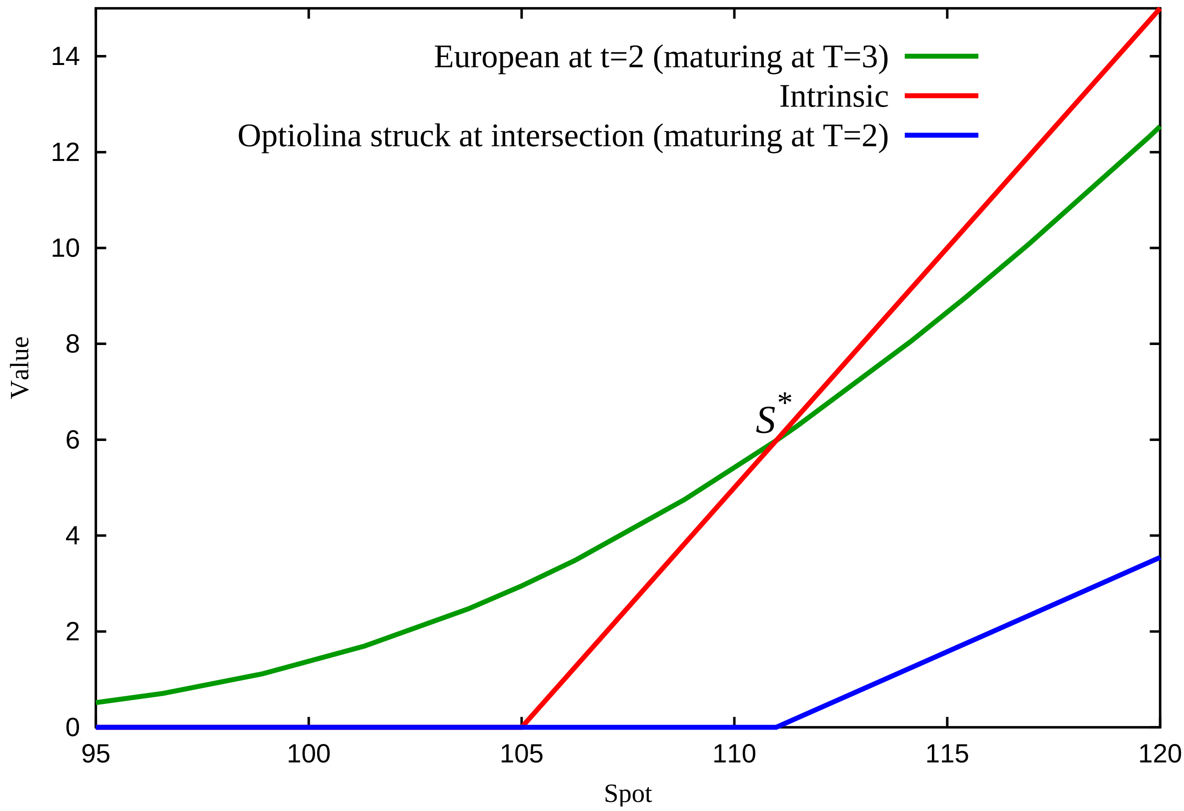

Figure 1: Value profile of a European-style call option struck at $K = 105$ and maturing at $T = 3$ as seen at $t = 2$, i.e. with one year to maturity. In this example, the domestic risk free short rate was assumed to be 4.25%, the continuous dividend yield (or risk free foreign short rate) was set at 6.5%, and volatility was fixed at 10% flat. It can be seen very clearly that for large S the value of the European call option behaves like the present value of the forward contract $F = S \cdot e^{-r_{\text{foreign}}(T-t)} - K \cdot e^{-r_{\text{domestic}}(T-t)}$. This means that the delta of the call option converges to a value less than one for increasing S. Also shown is the first adjustment option ("optiolina") in the backwards iterative procedure for static replication of American call options.

seen from the intersection point $\Delta(S^*_{t_{n-1}}, t_{n-1})$, we can amend the discounted expectation of the terminal profile beyond $S^*_{t_{n-1}}$ by adding another European option ("optiolina") struck and maturing just there, i.e. struck at $K_{t_{n-1}} := S^*_{t_{n-1}}$ and maturing at t_{n-1}. In order to make the profile of the now augmented portfolio look like that of the intrinsic value in the vicinity of $S^*_{t_{n-1}}$ we choose the notional h_{n-1} of the additional option to be[9]

$$h_{n-1} := 1 - \Delta(S^*_{t_{n-1}}, t_{n-1}) \ . \tag{18}$$

Since the delta of an American call option is exactly one on the exercise boundary, it will be typically very close to unity in our discrete approximation. This means that the notional for the second option to be added will typically be significantly smaller than that of the original contract. In particular, it will be the smaller, the finer the time discretisation is chosen to be. In figure 1, we show the first adjustment option of smaller notional for our example of a 3-year American call option with $n = 3$ steps in our time discretisation. Then, the resulting value profile of the portfolio satisfies the requirement to have the right pay-off at the final maturity t_n and never to be worth less than the intrinsic value at time t_{n-1} as can be seen in figure 2. At time t_{n-1}, one can therefore write

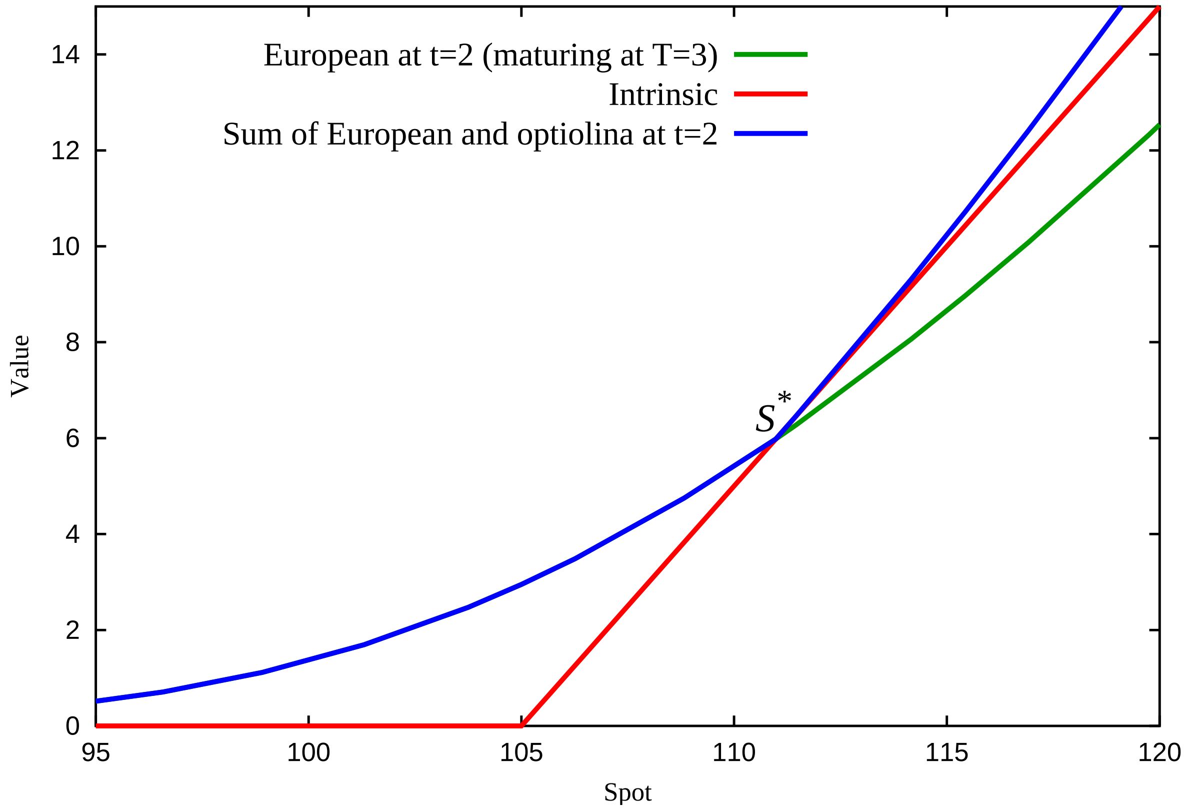

Figure 2: Value profile of the portfolio after adding the first adjustment option at time t_{n-1}. The parameters are the same as in figure 1.

$$\hat{C}_{\text{Am}}(S_{t_{n-1}}, t_{n-1}, K, T; n) = C_{\text{Eur}}(S_{t_{n-1}}, t_{n-1}, K, T) + h_{n-1} \cdot C_{\text{Eur}}(S_{t_{n-1}}, t_{n-1}, S^*_{t_{n-1}}, t_{n-1}) \ . \tag{19}$$

[9] Without loss of generality, we have assumed the notional of the American contract to be $h_n := 1$ for ease of notation since all of the optiolinas' notionals scale linearly in h_n.

Moving to time t_{n-2}, we see that the value of the three-times exercisable option is given by

$$\hat{C}_{\text{Am}}(S_{t_{n-2}}, t_{n-2}, K, t_n; n) = \max\left(\breve{\mathrm{E}}_{t_{n-2}}\left[\hat{C}_{\text{Am}}(S_{t_{n-1}}, t_{n-1}, K, t_n; n)\right], \max\left(S_{t_{n-2}} - K, 0\right)\right) \; . \tag{20}$$

Since, however, the value of the twice exercisable American option has been expressed as the sum of two European options, it is possible to calculate the value of this linear portfolio of European options (for which, once again, the Black-Scholes formula with the appropriate future implied volatility applies by definition).

The search procedure may now be repeated by first locating the intersection $S^*_{t_{n-2}}$ of the value of the portfolio with the intrinsic value of the American call option at time t_{n-2}. Note that the value profile of the portfolio and the intrinsic value as shown in figure 3 actually intersect twice. However, only the first intersection is of interest since the replication is only valid up to the approximated exercise boundary. In figures 4 and 5 we show the next optiolina to be added to the

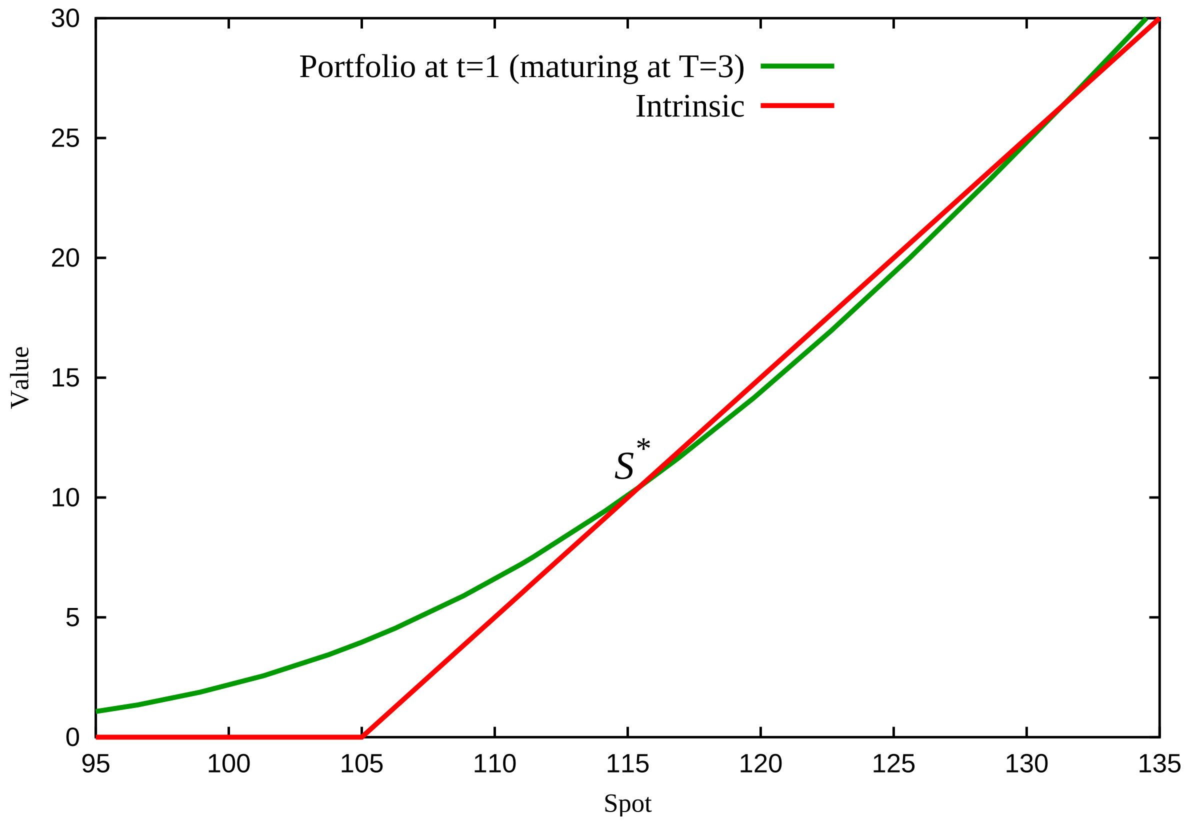

Figure 3: Value profile of the portfolio after adding the first adjustment option at time t_{n-2}. The parameters are the same as in figure 1.

portfolio, struck at $S^*_{t_{n-2}}$ and maturing at t_{n-2}, as well as the resulting portfolio profile as seen from t_{n-2}.

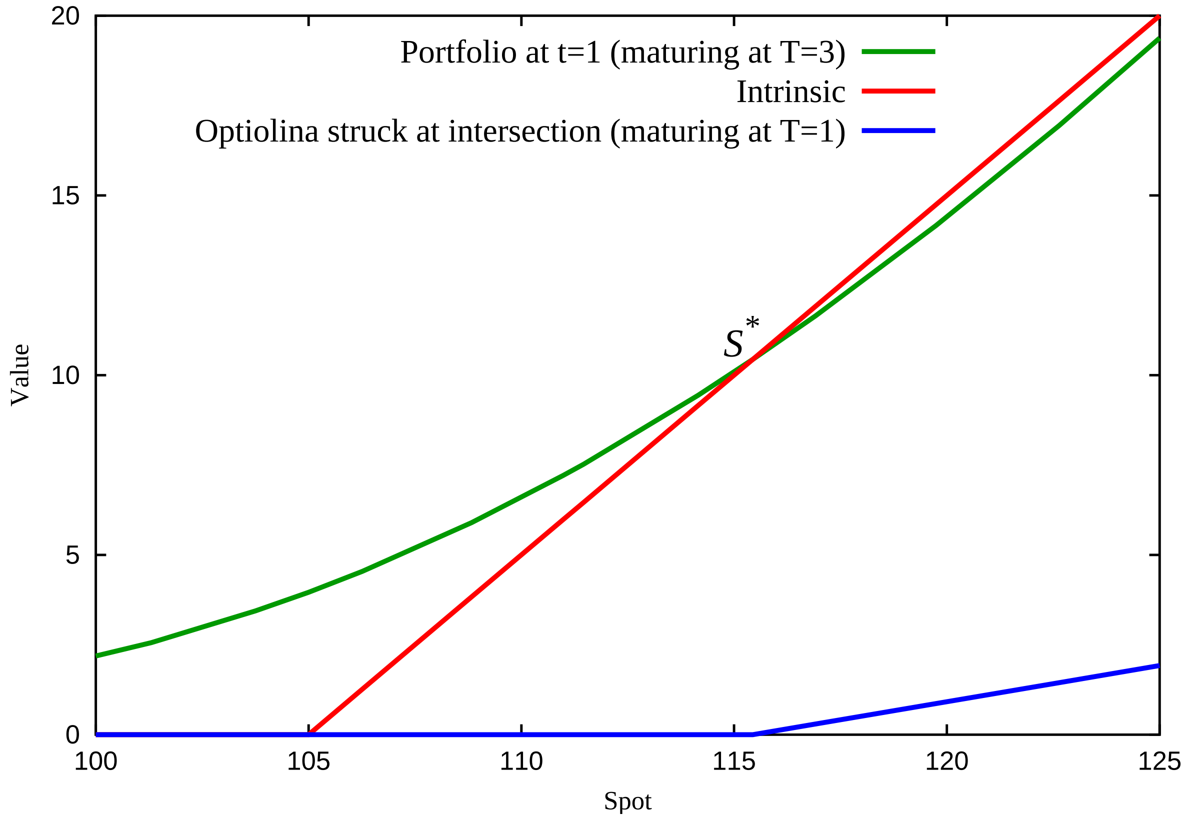

Figure 4: Value profile of the second adjustment option in the backwards iterative procedure for static replication of American call options. The parameters are the same as in figure 1.

It may also be noticed that the value profile of the portfolio actually exceeds that of the intrinsic value significantly for spot levels far above S^* in figures 2 and 5. This, however, is of no particular concern since the defining conditions for the replication are only that it reproduces the correct pay-off on the terminal maturity and on the exercise boundary and the correct value within those boundaries, and that it satisfies the associated von Neumann boundary condition, which is given in the limit of ever more time steps between inception and maturity.

The strategy employed therefore amounts to exploiting the fact that, at the exercise boundary, the value of the American option and of the intrinsic coincide, and by approximating the desired profile with a portfolio of appropriately chosen European options. For European options the Black-Scholes formula with the desired implied volatility then provides the solution, *irrespective of the process of the underlying*.

3.2 The rationale behind the method

As we shall show later on, the procedure described above can produce results of excellent numerical quality even when very few steps are used. This, we believe, is no accident, and can be

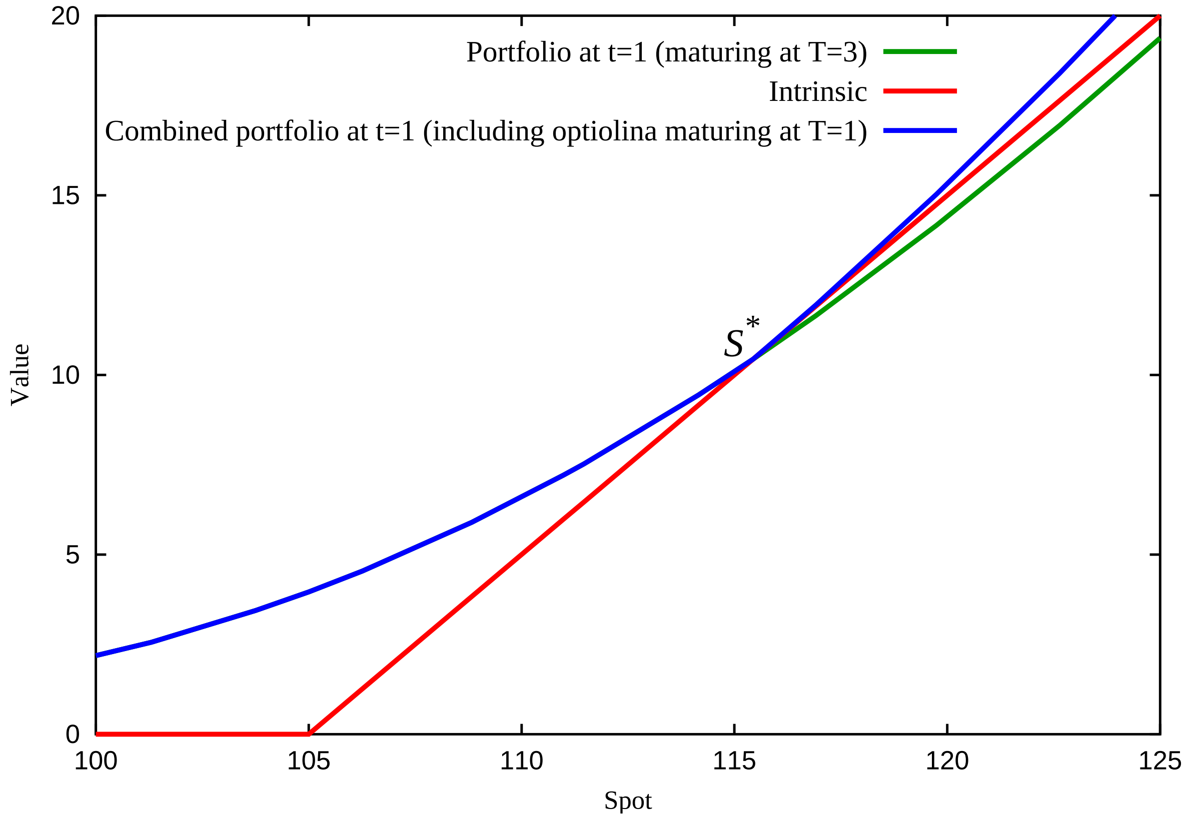

Figure 5: Value profile of the portfolio after adding the second adjustment option at time t_{n-2}. The parameters are the same as in figure 1.

explained by looking in very general terms at all the methods that use backward induction to locate the early-exercise boundary. In particular, all of these methods share the common feature that the problem is linear in its operand (which is the value of the American style option to be priced), i.e. the fact that it is theoretically possible to synthesize the present value of a European option of arbitrary terminal payoff profile by combining a suitable portfolio of plain-vanilla options which has the same intrinsic value at maturity. In common with conventional methods for the solution of the free boundary value problem for American-style options, the above-described procedure is an iterative method which moves backwards in time over a given time discretisation. At each step, an approximate solution to the governing partial differential equation is found on the earlier time slice t_i given that we know the value profile on a later time slice t_{i+1}. In all the backward induction methods, the additional constraints given by the American feature of the contract are then imposed[10] in order to ensure that the boundary conditions are met. From a formal point of view, methods such as finite difference techniques, trees, Fourier convolution approaches, etc., and,

[10] An exception is the *projected Successive Over Relaxation method* for finite differencing techniques [Wil98]. In practice, however,

formally, the new method presented here all can be seen as decomposing the value of the option, $\psi(S, t_i)$, into basis functions, $\phi_j(S, t_i)$:

$$\psi(S, t_i) = \sum_{j=0}^{n} a_{ij} \phi_j(S, t_i) \ , \tag{21}$$

where a_{ij} are the weights in the decomposition. In addition to the time discretisation, each of the approaches depends on another set of auxiliary discrete parameters required for the definition of the basis functions $\{\phi_j\}$ as listed in table 1. The individual functional shape of $\phi_j(S, t_i)$ depends

	Fourier convolution	Finite differencing	Replication method
Auxiliaries	Frequencies: $\{\omega_k\}$, $\omega_k = k\pi$, $k = 1 \ldots \frac{n-1}{2}$	Spot nodes: $\{S_k\}$, $k = 1 \ldots n$	Strikes & maturities of optiolinas: $\{(K_k, T_k)\}$, $k = 1 \ldots n$
$\phi_j(S, t_i)$	$\phi_0(S, t_i) = 1$ $\phi_{2k-1}(S, t_i) = \sin \omega_k S$ $\phi_{2k}(S, t_i) = \cos \omega_k S$	$\phi_k(S, t_i) = \begin{cases} 0 \text{ for } S < S_{k-1} \\ \frac{S-S_{k-1}}{S_k-S_{k-1}} \text{ for } S_{k-1} \leq S < S_k \\ \frac{S_{k+1}-S}{S_{k+1}-S_k} \text{ for } S_k \leq S < S_{k+1} \\ 0 \text{ for } S \geq S_{k+1} \end{cases}$	$\phi_k(S, t_i) = \begin{cases} C(S, t_i, K_k, T_k) \text{ for } t_i \leq T_k \\ 0 \text{ for } t_i > T_k \end{cases}$
a_{ij}	Weight of mode at time t_i	$a_{ij} = \psi(S_j, t_i)$	Notional of optiolina #j.

Table 1: A comparison of different discretisation methods for the numerical solution of the American option problem.

on the specific method. The basis functions would be, for instance, sines and cosines as a function of S for Fourier convolution techniques which account for modes in spot space. In the remainder of this section we shall concentrate, for the purpose of comparison, on the standard finite elements in spot space used for finite differencing methods (and thus for trees). The basis function set implicitly used by finite differencing methods consists of functions that are zero everywhere apart from the small interval between nodes[11] S_{j-1} and S_{j+1} where they change linearly from zero to one and back to zero. This choice of basis functions requires the solution of a tridiagonal linear system to derive the coefficients a_{ij} on time slice t_i from those on time slice t_{i+1}. In contrast, the approach suggested in this article utilises our knowledge of the solution of the problem for plain-vanilla options for the first of the two steps of the iterative procedure. Rather than approximating the solution to the boundary value problem by a discretisation method in spot space as well as in time, the solution is decomposed into basis functions *which all (approximately) solve the accompanying partial differential equation*. Since the governing equation is linear in its operand, a linear combination of solutions such as (21) is also a solution of the PDE. The focus is thus wholly placed on incorporating the American feature, i.e. we solve the free boundary value problem making maximum use of our understanding of the most closely related analytically soluble problem available (which are plain-vanilla European options). This, we believe, is the qualitative explanation for the excellent quality of the numerical results presented in the following section.

[11] In a finite difference method, each basis function $\phi_j(S, t)$ corresponds to a space discretisation node S_j.

4 Numerical results

In this section, we present the numerical values generated by the new method. Where possible, we also provide prices calculated using alternative methods for comparison.

4.1 Constant volatility

The first test any new methodology for the pricing of American style options must pass is to reproduce the value as given by conventional finite differencing methods for the standard Black-Scholes assumptions. In order to demonstrate the rapid convergence of the method as a function of the number of time slices used in the discretisation, we present in figure 6 the value for an American call option calculated by the replication method in comparison to the value as returned by the Barone-Adesi-Whaley approximation and a generalised Crank-Nicholson[12] solver of the standard Black-Scholes partial differential equation. The valuation parameters for this American call option were: Spot = 100, Strike = 105, Time to maturity = 2 years, domestic interest rate = 4.25% (continuosly compounded), foreign interest rate = 6.5% (continuosly compounded), annual volatility = 11.35%. For these settings, the price of an American call option is 2.88 and its vega is 0.47. For the PDE / Crank-Nicholson method, we used a specific transformation to logarithmic coordinates that removes the term containing the first spatial partial derivative. This term is known as the *convection* or *velocity* term and is primarily responsible for numerical instabilities [ZFV97]. By removing this term we achieve a remarkable level of robustness and rapid convergence for the finite difference method. This was done in order to provide a fair comparison of the finite difference method with the replication method. One should note that the stability-enhanced finite difference method is essentially converged for a grid of size 50×50. The same level of convergence, however, is reproduced by the replication method using only 6 time slices. To summarise the comparisons for constant volatility, we give some of the values from figure 6 in table 2, together with the delta, gamma, and vega statistics. Clearly, the method also works very well for the calculation of hedge ratios and risk statistics. It is also important to notice that, with very few replication steps, the numerical accuracy is orders of magnitude finer than the tightest bid/offer spread encountered in the market.

4.2 Term structure of implied volatility

Given a term structure of implied volatility,

$$\hat{\sigma}(T) = 10\% \cdot (1 + e^{-T}) \ ,$$

[12]This method is sometimes also referred to as the "temporal weighting finite difference method" [ZFV97] and is also known as the "θ-method" [Wil98]. The value of θ used was 0.55.

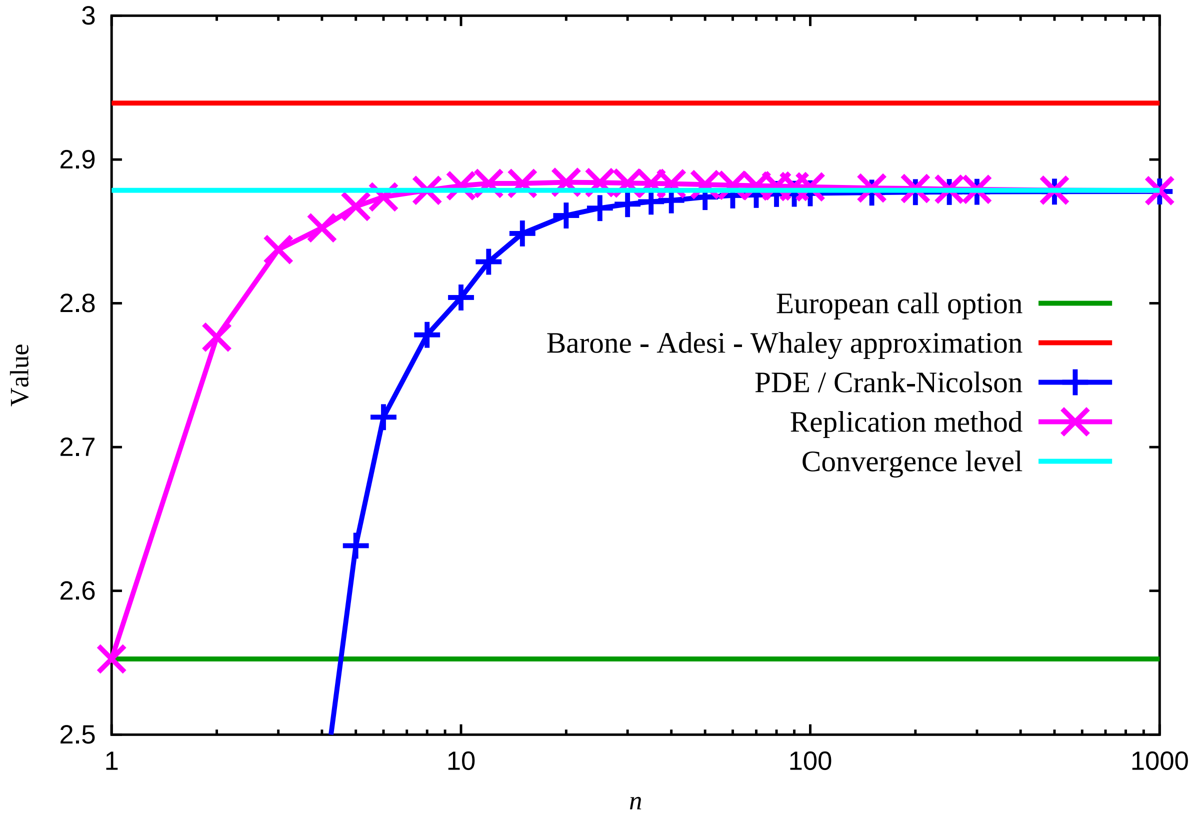

Figure 6: Convergence profile of the replication method for constant volatility as a function of the number of time slices, n, which is given on a logarithmic scale. Note that the replication value for $n = 1$ degenerates to the European call option value since then the only option in the replication portfolio is the one maturing on the final time slice. For the PDE method, a grid of $n \times n$ grid lines was used in the transformed coordinates (see text).

	Value	Delta	Gamma	Vega
European plain vanilla value	2.553	0.270	0.019	0.436
Barone-Adesi-Whaley approximation	2.939	0.314	0.024	0.475
2-step replication model	2.776	0.312	0.025	0.459
4-step replication model	2.852	0.315	0.025	0.467
8-step replication model	2.879	0.317	0.025	0.466
PDE model (501 spot levels x 507 time steps)	2.878	0.316	0.025	0.465

Table 2: Prices and greeks of conventional methods in comparison to the replication method for a flat volatility of 11.35%.

we carried out the same calculation as before. This term structure gives the same implied volatility for European options of time to maturity of 2 years, namely 11.35% as in the previous section. The results are shown in figure 7. For both the replication and the PDE method, future implied

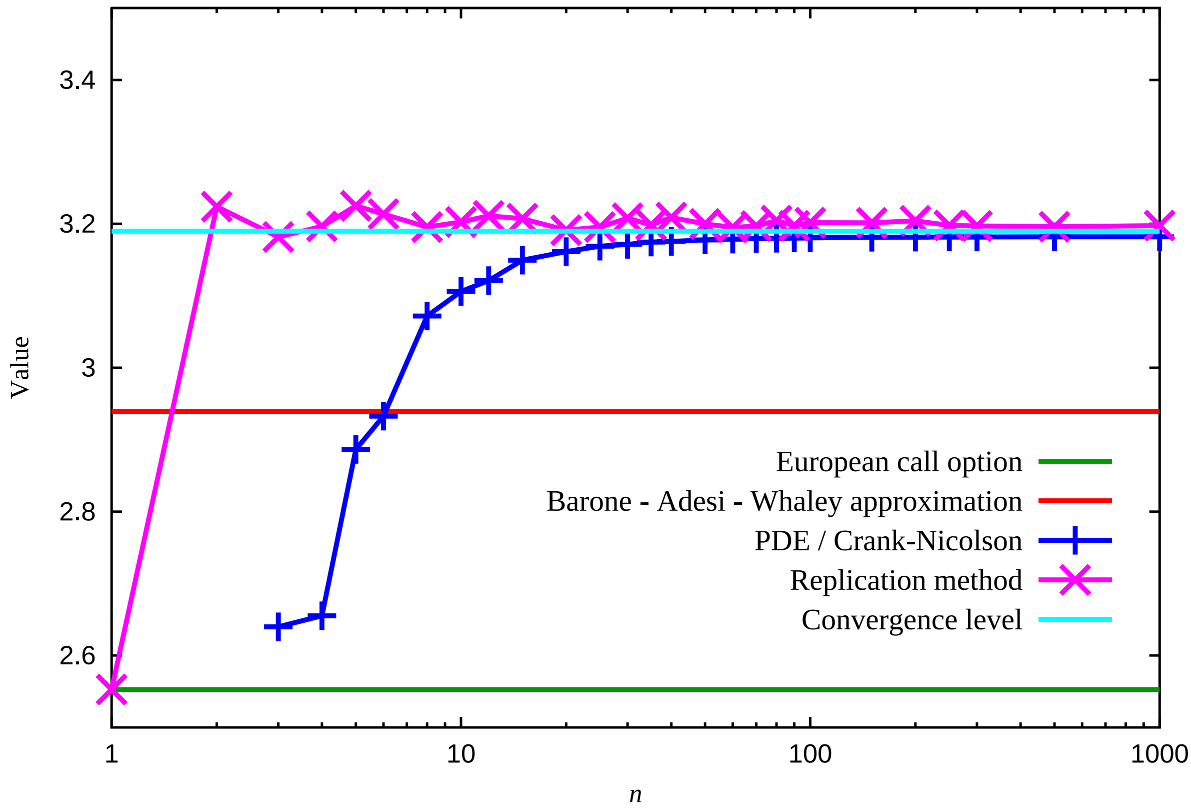

Figure 7: Convergence profile of the replication method for a term structure of volatility as a function of the number of time slices, n, which is given on a logarithmic scale. Note that the replication value for $n = 1$ degenerates to the European call option value since then the only option in the replication portfolio is the one maturing on the final time slice.

volatilities were calculated according to preservation of total variance, i.e.

$$\left[\hat{\sigma}(t_2,t_3)^2 \cdot (t_3 - t_2)\right] = \left[\hat{\sigma}(t_1,t_3)^2 \cdot (t_3 - t_1)\right] - \left[\hat{\sigma}(t_1,t_2)^2 \cdot (t_2 - t_1)\right] \quad (22)$$

for any $t_1 \leq t_2 \leq t_3$. Again, it can be seen that the replication method converges very rapidly to the value of 3.19, which, incidentally, is significantly above the value for the American call option when no term structure was taken into account. In table 3 we give the greeks produced by the replication method. The values given by the Barone-Adesi-Whaley method are only reproduced for completeness since they don't depend on the term structure of implied volatility.

	Value	Delta	Gamma	Vega
European plain vanilla value	2.553	0.270	0.019	0.436
Barone-Adesi-Whaley approximation	2.939	0.314	0.024	0.475
2-step replication model	3.224	0.361	0.028	0.436
4-step replication model	3.197	0.353	0.027	0.442
8-step replication model	3.196	0.350	0.026	0.448
PDE model (501 spot levels x 507 time steps)	3.182	0.347	0.026	0.448

Table 3: Prices and greeks of conventional methods in comparison to the replication method for a term structure of volatility ending in 11.35% according to $\hat{\sigma}(T) = 10\% \cdot (1 + e^{-T})$.

4.3 Future smile dynamics

The most demanding part of European and American option pricing is, arguably, the handling of the *future* implied volatility smile. As we stressed in the introductory sections, even a complete set of option prices today for a (double) continuum of strikes and maturities are not sufficient to pin down unambiguously the smile evolution. Implicitly or explicitly, it is therefore unavoidable to make a choice about this evolution whenever the spot implied volatility surface displays a strike dependence. Below we list a set of four simple possible ways of how the smile might evolve. In all cases, the implied volatility to be used to price an option when the current spot value of the underlying asset is S, the original spot value was S_0, the current calendar time is t, the maturity of the option is T, and its strike is K, is denoted by $\hat{\sigma}(S_0, S, t, K, T)$. The original implied volatility surface (for $t = 0$ and $S = S_0$) is given by $\hat{\sigma}_0(K, T)$.

- "Absolute floating": in this case, the future implied volatility is obtained from the original smile surface by simply reducing time to maturity and linearly offsetting the strike by how much the spot has moved, i.e.

$$\hat{\sigma}(S_0, S, t, K, T) = \hat{\sigma}_0(K + (S_0 - S), T - t) \; . \tag{23}$$

- "Absolute sticky": any dependence on the current spot level or calendar time is ignored, i.e.

$$\hat{\sigma}(S_0, S, t, K, T) = \hat{\sigma}_0(K, T) \; . \tag{24}$$

- "Relative floating": this is the closest one can get to what is otherwise referred to as "sticky delta" [Rei98]. The strike of an option is rescaled according to how the current spot evolved with respect to the spot at inception, i.e.

$$\hat{\sigma}(S_0, S, t, K, T) = \hat{\sigma}_0(K * \frac{S_0}{S}, T - t) \; . \tag{25}$$

- "Sticky strike": the initial volatility level is reduced according to preservation of total variance as it would be for a pure term structure model, and the volatility difference between at-the-money options and the required one is added, i.e.

$$\hat{\sigma}(S_0, S, t, K, T) = \sqrt{\frac{\hat{\sigma}_0(S,T)^2 \cdot T - \hat{\sigma}_0(S,t)^2 \cdot t}{T-t}} + \hat{\sigma}_0(K,T) - \hat{\sigma}_0(S,T)\ . \qquad (26)$$

This set is by no means intended to be comprehensive, it is simply meant to provide an indication of all the different behaviours for the smile one might conceive. Ultimately, the user will have to decide which of the above formulations, if any, satisfactorily describes the specific market at hand. However, it is of interest to see how these simple but plausible representations of smile dynamics already lead to significantly different prices of options.

In order to compare the values resulting from the replication method for these four possible future smile dynamics with a conventional method such as the Derman-Kani instantaneous volatility approach, we started from an assumed local volatility surface as given in figure 8. As can be

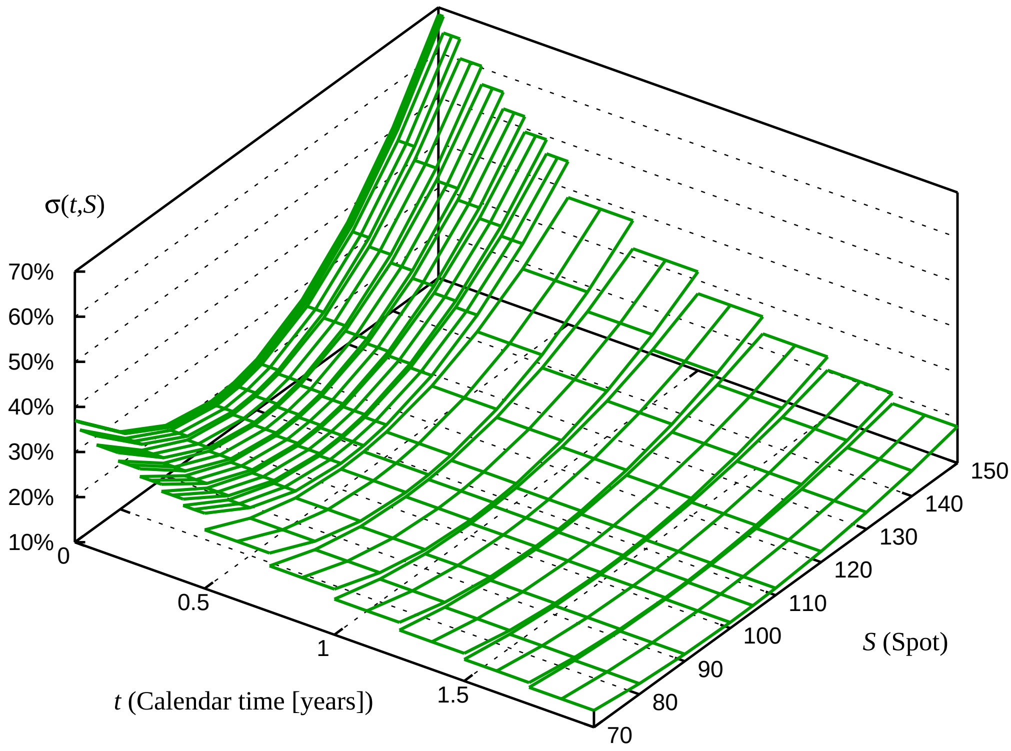

Figure 8: The instantaneous volatility surface used in the PDE method.

seen in the figure, the instantaneous volatility was assumed to be piecewise constant in calendar time. This, with the assumption of constant extrapolation wherever necessary, was converted to an implied volatility surface presented in figure 9 using the same PDE / finite differencing method mentioned in the previous sections. Using this initial implied volatility surface and keeping all

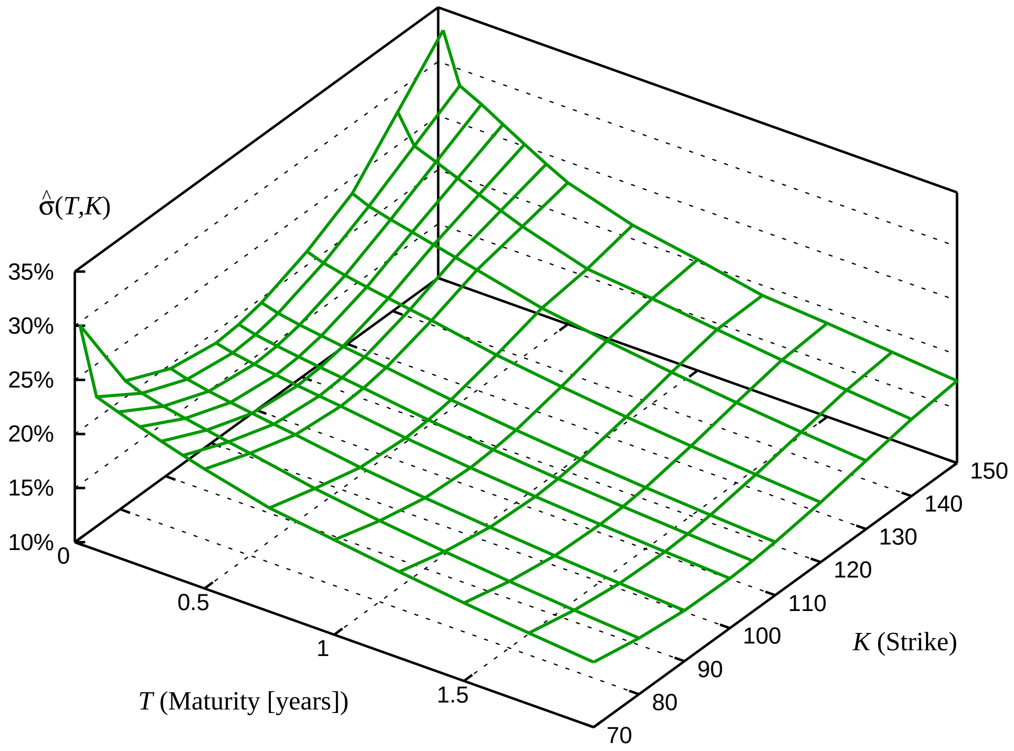

Figure 9: The initial implied volatility surface used in the replication method.

other parameters as before, we calculated the prices and greeks for all of the above suggested future smile dynamics. The results are presented in table 4. It should be stressed that the purpose of these calculations is to show the sensitivity of the American price to different smile assumptions and the speed of convergence in the presence of smiles. No agreement is, of course, to be expected between the various rows and the PDE results which have been obtained for a different future implied surface. As can be seen from the numbers, the variation in price due to a different assumption about the future smile dynamics is often between half and one vega. This gives a clear indication of the magnitude of the effect. Also in this more complex case, the convergence can still be observed to be rapid and robust, both for the price and for the greeks.

5 The exercise boundary

The boundary condition of an American style option is that on the exercise boundary its value becomes equal to the intrinsic value. In a continuous PDE formulation, this also implies that, for

Smile ending in 14.70% implied volatility at the strike	Value	Delta	Gamma	Vega
European plain vanilla value	4.069	0.320	0.016	0.466
Barone - Adesi - Whaley approximation	4.560	0.362	0.019	0.497
8-step replication model (absolute floating)	4.478	0.372	0.022	0.487
16-step replication model (absolute floating)	4.490	0.373	0.018	0.486
32-step replication model (absolute floating)	4.495	0.372	0.020	0.490
8-step replication model (absolute sticky)	4.609	0.374	0.020	0.498
16-step replication model (absolute sticky)	4.610	0.374	0.020	0.500
32-step replication model (absolute sticky)	4.609	0.374	0.020	0.499
8-step replication model (relative floating)	4.505	0.372	0.022	0.495
16-step replication model (relative floating)	4.514	0.372	0.022	0.497
32-step replication model (relative floating)	4.519	0.372	0.021	0.497
8-step replication model (sticky strike)	4.772	0.388	0.021	0.498
16-step replication model (sticky strike)	4.766	0.386	0.020	0.505
32-step replication model (sticky strike)	4.770	0.387	0.020	0.498
PDE model (2238 spot levels x 2013 time steps)	4.618	0.389	0.025	0.534

Table 4: Prices and greeks as given by the replication method for a pronounced smile of implied volatility. The European option value and the output of the Barone-Adesi-Whaley procedure are given solely for general comparison since they only use the implied volatility to maturity at the strike level. The PDE model values were calculated using the instantaneous volatility surface given in figure 8. Therefore, no agreement should be expected between the last and any of the rows above.

$t =$ constant, the delta must be continuous on the boundary [Wil98]. Since the delta is constant and of absolute value one on the outside, this means that the gamma, whilst in principle being slightly discontinuous on the boundary [ZS99], is typically of small value on the inside of the valuation domain near the exercise boundary. This in turn implies that the time value of an American option, i.e. its value minus its intrinsic value, increases slowly on the inside of the exercise boundary. It is by virtue of this initially slow increase of the time value of the American option value on the inside of the exercise boundary that a slight misplacement of the boundary (as it will occur for the replication method with very few steps) is of minor consequence for the resulting value of the option. In order to highlight this further, we propose the following definition. Let the region in Spot / calendar time space which is adjacent to the exercise boundary and in which the time value of the option is less than $q \cdot C_{\text{Am}}$ be called *the q-exercise band* of an American option (C_{Am} in this context is the value at inception). In figure 10, the location of the 10%-, the 5%-, and the 1%-exercise band is depicted as calculated from the PDE model for the smiley scenario that was used in section 4.3. Superimposed is the approximate exercise boundary as given by the absolutely sticky

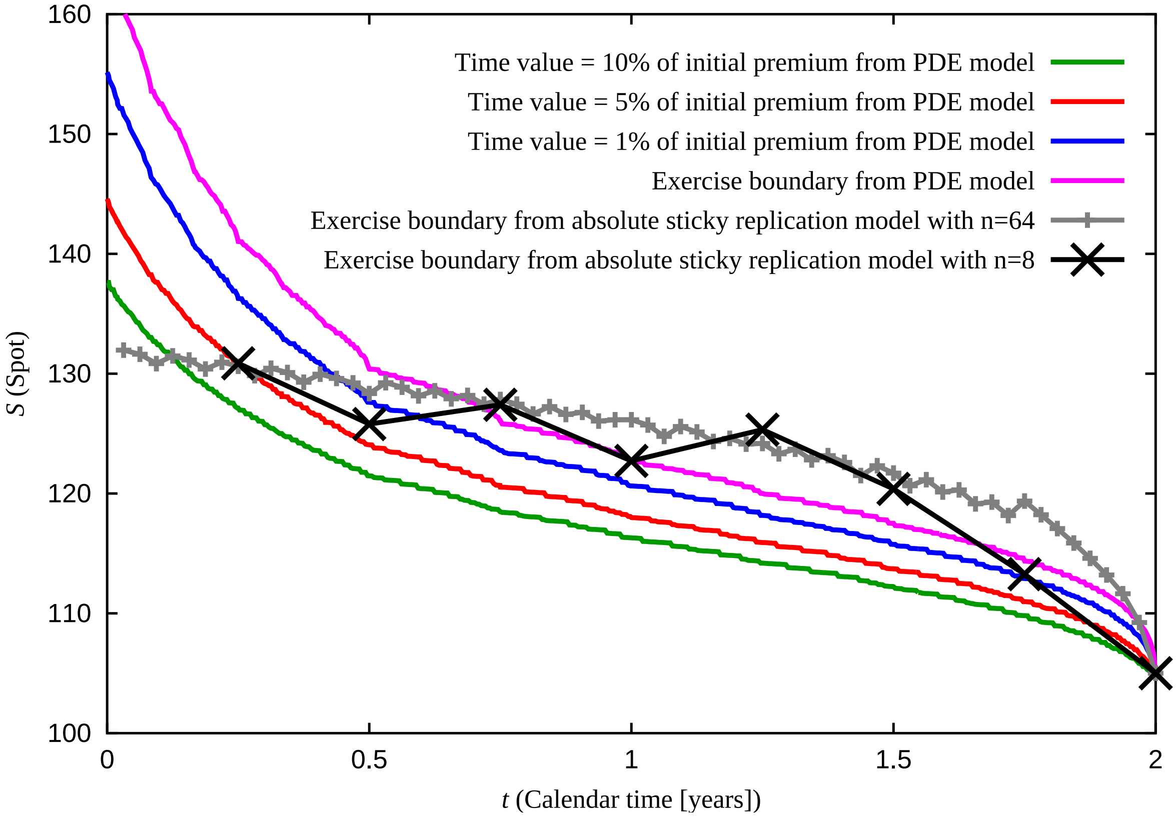

Figure 10: The 10%-, 5%, and 1% exercise bands and approximate exercise boundary for the absolute sticky replication model with $n = 8$.

replication model for $n = 8$ (absolutely sticky smile dynamics were chosen since these resulted in the value closest to the one given by the PDE model as can be seen in table 4). The observation that the approximate exercise boundary given by the replication model for a low number of replication times can be somewhat "jumpy" should therefore always be seen in the context of the fact that the resulting option value is actually very insensitive with respect to minor variations in the location of the exercise boundary.

As we have shown above, the minor inaccuracies in the approximate exercise boundary given by the replication model for small n have a very limited impact on to the value of the premium and the greeks. However, there is a simple way of improving the calculated exercise boundary approximation *without additional computational effort*. The replication method, in fact, does not require regular spacing of the replication times. Instead, any partitioning can be used. One may, for instance, from the observation of the typical shape of the exercise boundary or from analytical asymptotic expansions [Wil98], decide to use a square root scaling in order to cater for the natural square root shape of the exercise boundary near maturity. Clearly, the steeper the slope of the exercise boundary is the more will the replication method benefit from a more frequent rehedging. Thus, a better way to place the n hedge times may be to set them according to

$$t_i = T \cdot \left(1 - \left(1 - \frac{i}{n} \right)^2 \right) \quad . \tag{27}$$

In fact, as can be seen in figure 11, this simple scaling already has a dramatic smoothing effect on the approximate exercise boundary. An added bonus of the square-scaling of hedge times is also a

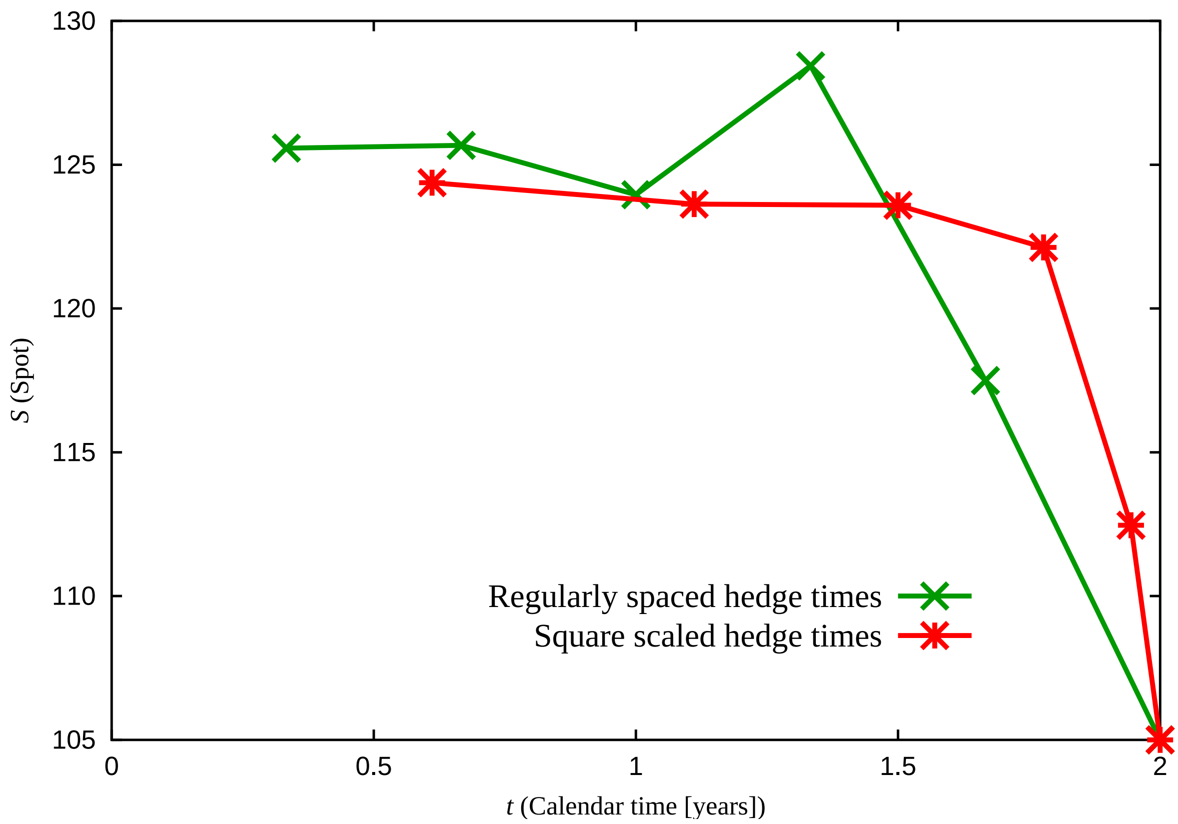

Figure 11: The smoothing of the approximate exercise boundary when hedge times are placed according to equation (27). The smile dynamics were assumed to be relatively floating for this diagram.

slightly improved convergence to the limit for $n \to \infty$ since replication corrections are then more frequent where they are most needed, i.e. where the exercise boundary has its steepest slope. This was also confirmed by our numerical observations. Finally, it is to be noted that scaling the hedge times and the resulting benefits of smoother exercise boundaries and increased convergence do not impose any additional computational effort on the replication method.

6 Conclusion

In this paper, we have proposed a computational method for the approximate but very accurate evaluation of American options. At the simplest level the method can be seen as a significant improvement in terms of computational efficiency over the asymptotically correct finite difference method. In this restricted form, the replication method here proposed can be applied to all those cases where it is assumed that current plain-vanilla option prices uniquely determine future volatilities, i.e. the case of constant or of purely time-dependent volatility. We claim, however, that the replication technique discussed in this paper can have a much wider range of applicability: whenever the current plain-vanilla option prices imply a strike dependence of the implied volatility, the evolution of the smile surface cannot be extracted without making strong assumptions about the process governing the evolution of the spot price. The more common and tractable assumptions, however, can predict future evolutions of the smile surface towards shapes at odds with past econometric experience and with trading intuition. Since we claim that the quality of any model should be judged on the basis of its ability to produce a plausible and desirable evolution of the smile surface, we extend a replication method that has already been used for *fixed* boundary problems to the case of American option. The approach directly leverages off the trader's intuition, and "talks his language" in terms of "sticky", "floating", "sticky-delta" smiles, etc.

It should also be noted that linearity of the governing evolution operator is no longer given if friction, i.e. transaction costs, is taken into account. Clearly, hedging a portfolio with partially offsetting positions will incur less total transaction costs than hedging all of the constituent contracts individually. However, it has been shown that a good approximation for the case of noticeable bid-offer spread is to use an appropriately increased volatility [Lel85, NK97, BV92, Wil98]. Also, the assumption of linearity is still justifiable in the presence of non-negligible bid-offer spreads if the portfolio only contains instruments of the same monotonicity of pay-offs, i.e. only plain-vanilla call options or only plain-vanilla put options of different strikes and maturities are used since then no offsetting effect will result. Luckily, this is the case when we replicate American options so that the method even applies when transaction costs are taken into account. The method proposed in this paper can therefore find some useful applications outside the domain of the perfect frictionless market assumptions.

References

[BS77] M. Brennan and E. Schwartz. The Valuation of American Put Options. *Journal of Finance*, 32:449–462, May 1977.

[BV92] P. Boyle and T. Vorst. Option replication in discrete time with transaction costs. *Journal of Finance*, 47:271, 1992.

[DKZ96] E. Derman, I. Kani, and J. Zou. The local volatility surface: Unlocking the information in option prices. *Financial analysts journal*, pages 25–36, July/August 1996.

[Ing98] J. E. Ingersoll. Approximating American options and other financial contracts using barrier derivatives. *Journal of Computational Finance*, 2(1), Fall 1998.

[Lel85] H. E. Leland. Option pricing and replication with transaction costs. *Journal of Finance*, 40:1283–1301, 1985.

[Mer73] R. C. Merton. Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, 4:141–183, Spring 1973.

[NK97] H. Neuhaus and Y. Kusuda. Pricing and Hedging Equity Options: The Effect of Market Structure and Transaction Costs. In *Equity Derivatives*. Risk Publications, November 1997.

[Par77] M. Parkinson. Option Pricing: The American Put. *Journal of Business*, 50:21–36, January 1977.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.

[Reb98] Riccardo Rebonato. *Interest rate option models*. Jon Wiley and Sons, 1998.

[Reb99] Riccardo Rebonato. *Volatility and Correlation*. Jon Wiley and Sons, 1999.

[Rei98] Eric Reiner. Understanding skew and smile behaviour in the context of jump processes and applying these results to the pricing and hedging of exotic options. In *Global Derivatives Conference 1998*, Paris, April 1998.

[Sch98] P. J. Schönbucher. A Market Model for Stochastic Implied Volatility. Discussion paper, Dept. of Statistics, Faculty of Economics, SFB 303, Universität Bonn, June 1998. ftp://ftp.wiwi.uni-bonn.de/papers/1999/b/bonnsfb453.pdf.

[Wil98] Paul Wilmott. *Derivatives*. Jon Wiley and Sons, 1998.

[ZFV97] R. Zvan, P. A. Forsyth, and K. R. Vetzal. Robust numerical methods for PDE models of Asian options. *Journal of Computational Finance*, 1(2):39–78, Winter 1997.

[ZS99] You-Ian Zhu and Yingjun Sun. The singularity-separating method for two-factor convertible bonds. *Journal of Computational Finance*, 3(1):91–110, Fall 1999.