The Logistic Function and Implied Volatility: Quadratic Approximation and Beyond

Etienne Hofstetter[a] and Michaël J. P. Selby[a,b,1]

[a]Centre for Quantitative Finance, Imperial College, London SW7 2BX, United Kingdom
[b]Financial Options Research Centre, Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom

11 January 2001

[1]Corresponding author. Tel.: +44-20-7289-1015; fax: +44-20-7581-8809; e-mail: michaeljpselby@beeb.net
The Authors thank Ian Buckley, Simon Babbs and Robert Tompkins for useful comments.

The Logistic Function and Implied Volatility: Quadratic Approximation and Beyond

Abstract

We introduce a new methodology for estimating implied volatilities, and other option pricing parameters. Almost all valuation formulae are linear combinations of the special functions, whose arguments contain some or all of the parameters. We obtain our estimates by replacing these functions with a surrogate. Consequently, we obtain *simple* formulae, when options are not necessarily at-the-money. For the extended Black-Scholes-Merton formula, the logistic distribution replaces the cumulative normal distribution. These formulae, which are identical for both European puts and calls, are at least quadratic approximations, and substantially extend and improve previous approximation validity ranges.

JEL classification:
*Key words:*Approximation, Extended Black-Scholes-Merton Formula, Exponential Sums, Implied Volatility, Logistic Distribution, Option Pricing, Rational Approximation, Solution by Radicals.

I Introduction

Implied volatilities[2] play a virtually fundamental role in option pricing, trading and hedging. Further, the use and determination of implied parameter values have become increasingly important in empirical derivatives research[3]. Almost all contingent claims valuation formulae are linear combinations of the special functions, with arguments containing most, if not all, of the parameters[4]. Even when approximations are made directly to the special functions, it is very difficult, if not impossible, to produce a "simple" parameter inversion formula that is based solely on elementary functions (or even special functions). Our prime objective is to find simple approximations for determining values of implied parameters, where we define a simple approximation to be one obtained as an algebraic solution, or equivalently, a solution by radicals[5]. Simple solutions are desirous because they have two very attractive properties. They are easy to implement, and provide very fast computational algorithms. Motivated by computational efficiency, this paper focuses on deriving an approximation algorithm for determining the implied volatility of extended Black-Scholes-Merton (BSM) put and call options. It is robust to the moneyness of the option, requires few operations, uses only elementary functions, and is the same for both European put and call options. It is sufficiently easy to implement that, at most, only a pocket calculator is needed.

Brenner and Subrahmanyam (1988, 1994)[6] developed approximations for the BSM implied parameters, based on Laplace's (1785) power series expansion of the cumulative normal distribution function (cndf). They argued that approximations should be based on the assumption that the stock (underlying) price is equal to the discounted exercise price, as in practice much options business is transacted around the at-the-money position. This assumption is by no means innocuous. When geometric Brownian motion describes the dynamics of the underlying, the argument of the cndf significantly simplifies. This enabled Brenner and Subrahmanyam to essentially use a linear approximation to obtain their formula for estimating the at-the-money implied volatility, which, within a rather limited range, proved to be quite accurate. However, it is rare for transactions to actually take place exactly at-the-money and, as we shall show, the moneyness of the option position significantly affects the efficiency of their (and others') approximations[7]. Further, their fundamental assumption does not necessarily help when using other pricing models based upon different stochastic processes.

---

[2]It is both important and interesting to note that Merton (1973), page 161, stated that the Black-Scholes formula "does depend on the rate of interest (an "observable") and the total variance of the return on the common stock which is often a stable number and hence, accurate estimates are possible from time series data". In Merton (1990), pp. 282-283, footnote 27, this statement was amended by reference to the fact that "subsequently, several researchers have reversed this process and have used option prices and the model to deduce implied variance rates as estimates for future volatility of the stock".

[3]For example, in research on the relationship between options and their associated underlying asset markets, the implied value of the underlying and the volatility must be jointly determined. See, for example, Stephan and Whaley (1990), Snelling (1987), and Varson and Selby (1997).

[4]Plain vanilla, infinitely lived, American put and call options, subject to a continuous dividend stream, are rare examples of contracts whose valuation formulae do not use special functions. See Merton (1973).

[5]These solutions must be obtained by only a finite number of additions, multiplications, subtractions, divisions, and extraction of roots. See Bell (1937), page 340.

[6]See also Bharadia et al. (1996), Corrado and Miller (1996), and Chance (1996).

[7]We are not the first to make this point. See for example Bharadia et al. (1996), Corrado and Miller (1996) and, Chance (1996).

In contrast with previous research, we obtain our approximations by replacing the special function itself. Focusing on the extended BSM formula[8], the cndf is replaced by the logistic distribution. As a result, for any implied parameter, we are able to obtain simple formulae, which are valid when options are not necessarily at-the-money, as well as encompassing all previous results as special cases. These formulae are not computationally intensive. They are second order approximations and robust to moneyness, thus allowing us to substantially extend and improve the validity of the approximation range. Further, our approach is extendible to most other valuation formulae, as it is essentially a rational approximation by exponential sums.

The remainder of the paper is organized as follows. In Section II we define our notation and terminology. In Section III we briefly describe the logistic distribution, paying particular attention, as an approximating function, to its relationship with the cumulative normal distribution. In Section IV we describe our methodology and apply it to derive new basic formulae for the implied volatility of a European option. We also compare and contrast our results with previous work, thus motivating our final formula. In Section V we present our summary and conclusions.

II Notation and Terminology

Following Black and Scholes (1973), and Merton (1973, 1974) we define for European style stock options on underlying assets paying (continuous) dividends:

C = the price of a European call option on a stock,
C_A = the approximated price of a European call option on a stock,
P = the price of a European put option on a stock,
P_A = the approximated price of a European put option on a stock,
S = the stock price,
X = the option strike price,
r = the continuously compounded riskless rate of interest,
a = the continuous dividend rate, proportional to the value of the stock,
t = the time to maturity in years,
σ = the instantaneous standard deviation of returns on the stock,
σ_i = the implied instantaneous standard deviation of returns on the stock,
$N(\cdot)$ = the cumulative standard normal distribution function (cndf),
$n(\cdot)$ = the standard normal probability density function,
$N_A(\cdot)$ = the logistic function
$d = X\exp(-rt)/(S\exp(-at))$, (moneyness ratio[9]),
$\alpha = -\ln(d)/(\sigma\sqrt{t})$,
$C^* = C/(S\exp(-at))$,
$C^*_A = C_A/(S\exp(-at))$,

[8]The extended BSM formula is frequently called the Garman-Kohlhagen formula. However, it essentially appears as Footnote 62 in Merton (1973) (with a typographical error). It is also contained within equation (8) of Black and Cox (1976), and can thus be directly obtained from that formula.

[9]The symbol d (with $a = 0$) was introduced in Merton (1974) as the "quasi-debt ratio". This parameter is of fundamental importance in the current analysis. Although in this paper it might be felt by some that $1/d$ is a more natural parameter, we retain the Merton (1974) definition rather than introduce additional notation.

$C^{**} = C/(X\exp(-rt))$,
$P^* = P/(S\exp(-at))$,
$P^*_A = P_A/(S\exp(-at))$,
$P^{**} = P/(X\exp(-rt))$.

The values of European call and put options, C and P, are given by the formulae

$$C = S\exp(-at)N(d_1) - X\exp(-rt)N(d_2), \tag{1}$$

$$P = X\exp(-rt)N(-d_2) - S\exp(-at)N(-d_1), \tag{2}$$

where $d_1 = d_2 + \sigma\sqrt{t}$, and

$$d_1 = -\frac{\ln(d)}{\sigma\sqrt{t}} + \frac{1}{2}\sigma\sqrt{t} = \alpha + \frac{1}{2}\sigma\sqrt{t}. \tag{3}$$

The prime objective of Black and Scholes was the determination of the market values of options given S, X r, a, t and σ. However, early empirical studies of their valuation formulae and extended forms, demonstrated that they did not describe actual prices well. Although there are many reasons why this was not totally unexpected, the formulae have a variety of attractive properties and there is, therefore, a great reluctance to discard such elegant results without first considering all aspects of the assumptions, for example, the strong one of lognormality of returns. However, whatever assumptions one makes in regard to stochastic processes, in any model the one parameter that cannot be directly observed is σ, and much effort has been directed at this point[10]. Since at least the publication of Latané and Rendleman (1976) great emphasis has been placed on both the use and the determination of implied volatilities, σ_i. As a result, σ_i given S, X, r, a, t and C or P, must be determined using either equation (1) or (2), although there is no elementary method for the inversion of these formulae. Consequently, all estimates of σ_i are based on numerical approximation and/or iterative/search procedures, and normally require the use of a computer. Although iterative procedures are mathematically attractive, they are frequently computationally intensive. Numerical approximation, however, is usually based on polynomial algorithms, which if well designed, is both fast and accurate over the relevant approximation region.

There are two principle difficulties in deriving formulae for BSM implied parameters. First, to obtain a suitable proxy inverse function, the BSM formula itself must be approximated. The natural approach is to replace the cndf by a power series. However, such series do not generally converge quickly[11]. This is particularly true of Laplace's (1785) series, which is the basis of the approach of Brenner and Subrahmanyam (1988, 1994). Second, power series arising directly from the cndf are of odd order. This is of particular importance for implied volatility estimation when the option is not at-the-money. Substituting d_1 and d_2 into

[10] We assume that an explicit price of risk function is not required, as is the case with the Black and Scholes (1973), and Merton (1973) models. Further, Merton clearly demonstrated that even when σ is indeed constant, one still needs the expected future total volatility over the remaining life of the contract, the essential functional form of the formulae remaining unchanged.

[11] Rational approximations based on such series are computationally attractive, but they may exhibit very bad behaviour outside their validity range. Further, they not do not lead to facile formulae, and are also difficult to remember!

the power series leads to first order approximations that require the solution of a quadratic equation, which is readily solvable. However, second order approximations, because of the odd order property of the series, lead to a sextic equation. It is a classic result that the roots of quintic and higher order polynomial equations cannot usually be obtained as "a solution by radicals". Therefore, in contrast with first order approximations, it is impossible to obtain an easy to use general formula for the implied volatility.

Dividing (1) and (2) throughout by $S\exp(-at)$ and $X\exp(-rt)$, respectively, we obtain our prime representations for analysis and approximation:

$$C^* = N(d_1) - dN(d_2), \tag{4}$$

$$P^* = dN(-d_2) - N(-d_1), \tag{5}$$

$$C^{**} = \frac{1}{d}N(d_1) - N(d_2), \quad \text{and} \tag{6}$$

$$P^{**} = N(-d_2) - \frac{1}{d}N(-d_1). \tag{7}$$

Merton (1973) proved[12] that the BSM formula is homogeneous of degree one in both $S\exp(-at)$ and $X\exp(-rt)$, and that the two ratios $C^* = C/(S\exp(-at))$ and $C^{**} = C/(X\exp(-rt))$, and their equivalent for puts, are strictly two parameter functions, the parameters being d and $\sigma^2 t$. These ratios represent a switch of numeraire[13] (see Johnson (1987)). Although the market presents both underlying and strike prices, for the purpose of this paper it is highly convenient to work with the moneyness ratio, d. Our approach enables us to focus on the prime properties of the formulae. When BSM options are at-the-money[14], $d = 1$, the arguments d_1 and d_2 are then pure functions of the total expected future volatility of the stock, $\sigma^2 t$.

Finally, in passing, we note that we can write the Put-Call Parity Theorem relationship in the form[15]:

$$P^* = C^* + (d - 1), \tag{8}$$

or

$$P^{**} = C^{**} + \frac{(d-1)}{d}. \tag{9}$$

Clearly, through the use of put-call parity, we can obtain put parameter approximations (such as implied volatility) directly from those of calls. However, from both a practical view and one of error analysis, it is superior to give approximations directly related to P. Finally, we shall only work with the C^* and P^* formulae, because when $a = 0$ the resulting approximations are more accurate.

[12]Pp. 165-6. Further, under suitable conditions Merton proved, Theorem 9 (page 149), that options, in general, are homogeneous of degree one in both the underlying and the strike price.

[13]It is interesting to note that we are working within the spirit of Samuelson (1965) who suggested that option prices should be quoted in units per strike price.

[14]In a trading environment the phrase "at-the-money" means that the price of the underlying is exactly equal to the strike price. However, in option theory, it is now common to use our definition because when $d = 1$ the logarithmic component in d_1 and d_2, is eliminated. Our definition is essentially "at-the-money on a forward basis"; see Brenner and Subrahmanyam (1994), page 26.

[15]We note that $dP^{**} = P^*$

III Logistic Function Approximation

In the previous section we emphasized the need to approximate the BSM formula and that power series approximation is not ideal. However, for about 60 years it has been known that the logistic distribution, with suitable parameter values, approximates the cndf well over a large range. Indeed Finney (1978), p. 47, states that the two models

$$U = \frac{D}{1 + \exp\{-2(\alpha + \beta x)\}} \tag{10}$$

and

$$U = D \int_{-\infty}^{\alpha+\beta x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) dt \tag{11}$$

are practically indistinguishable on empirical evidence. This observation has been used to develop a variety of approximations to the cndf[16]. To achieve our desire to find *simple* approximations we use an explicitly designed logistic function as a surrogate for the cdnf, and thus replace the cdnf everywhere in the BSM formula. We shall show that this leads to a new power series expansion having more desirable properties than those of the direct normal expansions. Accordingly, in this section we derive and describe the form of the logistic that we shall use as the basis of our approximations.

The cndf, $N(x)$, can be written as follows:

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) dt = \frac{1}{2} + \int_{0}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}t^2) dt. \tag{12}$$

Therefore, when approximating $N(x)$ it is only necessary to focus on the integral over the finite range[17] $[0, x]$. We now introduce the following version of the logistic function, denoted by $N_A(x)$, which will be our approximating function, once β has been determined:

$$N_A(x) = \frac{1}{1 + \exp\{-\beta x\}} = \frac{1}{2}[1 + \tanh(\frac{\beta x}{2})] \ . \tag{13}$$

First, we note that $N(x)$ and $N_A(x)$ have the same asymptotic behaviour for $x \to \pm\infty$ and the same value for $x = 0$, they are both symmetric about this point, and both functions are concave for all $x \leq 0$ and convex for all $x \geq 0$. Frequently, in Statistics, β is chosen[18] so as to match the first two moments of the logistic function with the corresponding ones from the cdnf, which gives a value of $\beta = \pi/\sqrt{3}$. In contrast, we propose an alternative approach that recognizes the quasi-linear behaviour of the two functions around $x = 0$, and also the importance of near-at-the-money positions. We match[19] the slopes of both functions at $x = 0$ and obtain $\beta = \sqrt{8/\pi}$, which gives a significantly improved estimator of $N(x)$, as measured by Johnson et al. (1995).

[16] See the references in Johnson et al. (1994), p. 118 and, Johnson et al. (1995) p.119-120

[17] We do not assume that $x \geq 0$ necessarily.

[18] See Johnson et al. (1995).

[19] This approach is much more in the spirit of approximation theory and practice. See, for example, spline functions.

Now the Laplace series expansion for $N(x)$ is

$$N(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} (-1)^{-n} \frac{x^{2n+1}}{[n!2^n(2n+1)]} \; , \tag{14}$$

therefore,

$$N(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(x - \frac{x^3}{6} + \frac{x^5}{40} - \ldots \right) \; . \tag{15}$$

The Taylor series for $\tanh(x)$ is

$$\tanh(x) = \sum_{n=0}^{\infty} \left(2^{2n}(2^{2n} - 1)B_{2n} \frac{x^{2n-1}}{(2n)!} \right) \qquad (|x| < \frac{\pi}{2}) \tag{16}$$

where B_n is the nth Bernoulli number, and therefore,

$$\tanh x = x - (\frac{x^3}{3}) + (\frac{2x^5}{15}) - (\frac{17x^7}{315}) + \ldots + [\frac{2^{2n}(2^{2n}2n - 1)B_{2n}x^{2n-1}}{(2n)!}] + \ldots \tag{17}$$

Substituting $\beta = \sqrt{(8/\pi)}$ into (13) and using (17), we obtain

$$N_A(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(x - \frac{2x^3}{3\pi} + \frac{8x^5}{15\pi^2} - \ldots \right) \; . \tag{18}$$

If one views the logistic function as a rational approximation of the cndf, the argument, x, being rescaled via a negative exponential transformation reflecting the cndf's rapid convergence to its asymptotes, we can see why it is such a good surogate for the cndf. Comparing Eq. (18) and Eq. (15) we see that both functions have the same form of algebraic series expansion, differing only in the coefficients of x^n for $n > 1$. However, the small difference between each coefficient is crucial. Whereas the series expansion of $N(x)$ cannot be expressed as a simple function, that of $N_A(x)$ obviously can be. For implied parameter, and especially implied volatility estimation, this is of crucial importance when the option is not at-the-money. In the next section we shall show that in this case, substituting d_1 and d_2 into a power series based on the logistic function leads to higher order approximations that require the solution only of a quadratic equation.

A Put-Call Parity and the Logistic Function

The Put-Call Parity Theorem states that

$$S\exp(-at) + P = C + X\exp(-rt), \tag{19}$$

and because $N(x) = 1 - N(-x)$ it is easy to prove that the BSM put and call pricing formulae [(1) and (2),respectively] satisfy this relationship. When $N(x)$ is replaced in these formulae by $N_A(x) = 1/[1 + \exp(-\beta x)]$ it is not immediately obvious that the resulting formulae will necessarily satisfy (19). We shall now prove that C_A and P_A do in fact satisfy (19), and that this is the key to understanding why our results are independent of the style of the options.

Theorem 1 *Let $M(x)$ represent the cumulative distribution function of a symmetric random variable representing the underlying asset, and let $M_c(x)[= 1 - M(x) = M(-x)]$ represent the complementary distribution function. Then the Put-Call Parity Theorem is satisfied if the simple European call and put pricing formulae have the form*
$C = S\exp(-at)M(x_1) - X\exp(-rt)M(x_2)$
and
$P = X\exp(-rt)M_c(x_2) - S\exp(-at)M_c(x_1)$.

Proof: Consider

$$\begin{aligned} C - P = &\ S\exp(-at)[M(x_1) + M_c(x_1)] - X\exp(-rt)[M(x_2) + M_c(x_2)] \\ = &\ S\exp(-at) - X\exp(-rt). \end{aligned} \tag{20}$$

Now from Margarbe (1978) and Johnson (1987) we see that if C is a call option,then by the change of numeraire technique, P must be the associated put option written on identical terms. Hence, the required result immediately follows □.

Turning now to the logistic function, $N_A(x)$, which satisfies the conditions of the theorem, we have:

$$C_A^* = \frac{1}{1+\exp(-\beta d_1)} - \frac{d}{1+\exp(-\beta d_2)} \tag{21}$$

and

$$P_A^* = \frac{d}{1+\exp(\beta d_2)} - \frac{1}{1+\exp(\beta d_1)}. \tag{22}$$

Thus

$$\begin{aligned} C_A^* - P_a^* = &\frac{1}{[1+\exp(-\beta d_1)]} - \frac{d}{[1+\exp(-\beta d_2)]} - \\ &\frac{d\exp(-\beta d_2)}{[1+\exp(-\beta d_2)]} + \frac{\exp(-\beta d_1)}{[1+\exp(-\beta d_1)]} = 1 - d, \end{aligned} \tag{23}$$

which is expected.

IV European Option Implied Volatility Estimaton

A Derivation of Approximation Formulae

Further, much traded options business is concentrated within a rather small region, ±10% of the at-the-money position. Brenner and Subrahmanyam (1988) gave a simple and accurate formula for approximating the at-the-money implied volatility, σ_i. However, even minor variations from $d = 1$ lead to poor results. Subsequently, for positions when options are not necessarily at-the-money, Bharadia et al. (1996), and Corrado and Miller (1996) directly extended Brenner and Subrahmanyam's results. Bharadia et al. (1996), and Corrado and Miller (1996) derived a quadratic equation that must be satisfied by the implied volatility.

However, their equation is fundamentally based on a linear approximation to the cndf and, therefore, the accuracy of their estimates is limited to a relatively small range around the at-the-money position. Following a different path Chance (1996), taking the Brenner and Subrahmanyam estimate as a starting position, "adds terms to reflect the option's moneyness and sensitivity to a difference in the standard deviation", the prime adjustment term also arising from the solution of a quadratic equation. This method is computationally more intensive than the others, and reflects the fact that the Brenner and Subrahmanyam estimate is by no means an optimal choice of starting position.

In this section we shall show that our results contain all previous results. We improve upon them, whenever it is possible, with no increase in computational intensity or complexity by deriving new formulae for approximating the implied volatility of a European option, which is not necessarily at-the-money. We shall achieve our objective by replacing the cndf in the BSM formula by the logistic function and undertake a power series expansion. We shall see that this leads to polynomials of all orders in the implied volatility and, therefore, solution by radicals algorithms will be obtained.

The approximated European call option price, obtained from (4), is:

$$C_A^* = N_A(d_1) - dN_A(d_2). \tag{24}$$

In Figures 1 and 2 we plot C and C_A for the maturities $t = 1/4$ and $t = 1/12$, respectively. We see that (24) is a very good approximation to (4) and, consequently, it is reasonable to expect that we can obtain a finite power series, based on (24), which will approximate (4) well. Therefore, undertaking a Taylor's series expansion of $N_A(d_1)$ and $N_A(d_2)$ about $d = 1$, emphasizing the functional dependence on d by writing $d_1 = d_1(d)$ and $d_2 = d_2(d)$, and substituting the two series into (24) we obtain:

$$\begin{aligned} C_A^* &= N_A(d_1(1)) + (d_1(d) - d_1(1))N_A'(d_1(1)) + \frac{1}{2}(d_1(d) - d_1(1))^2 N_A''(d_1(1)) + \dots \\ &\quad -d\left[N_A(d_2(1)) + (d_2(d) - d_2(1))N_A'(d_2(1)) + \frac{1}{2}(d_2(d) - d_2(1))^2 N_A''(d_2(1)) + \dots\right]. \end{aligned} \tag{25}$$

Remembering that we defined $\alpha \equiv d_1(d) - d_1(1) = d_2(d) - d_2(1) = -\ln(d)/(\sigma\sqrt{t})$ we can write (25) as follows[20]:

$$\begin{aligned} C_A^* = &\quad N_A(d_1(1)) + \alpha N_A'(d_1(1)) + \tfrac{1}{2}\alpha^2 N_A''(d_1(1)) + \dots \\ &-d\left[N_A(d_2(1)) + \alpha N_A'(d_2(1)) + \tfrac{1}{2}\alpha^2 N_A''(d_2(1)) + \dots\right] \\ = &\ [N_A(d_1(1)) - dN_A(d_2(1))] + \alpha\,[N_A'(d_1(1)) - dN_A'(d_2(1))] + \\ &\ \tfrac{1}{2}\alpha^2\,[N_A''(d_1(1)) - dN_A''(d_2(1))] + \dots \end{aligned} \tag{26}$$

where

$$N_A(x) = \frac{1}{1 + \exp(-\beta x)} \tag{27}$$

[20]There exists a similar expression for (4).

$$N'_A(x) = \beta \exp(-\beta x) \left(\frac{1}{1+\exp(-\beta x)} \right)^2 \tag{28}$$

$$N''_A(x) = -\beta^2 \exp(-\beta x)(1-\exp(-\beta x)) \left(\frac{1}{1+\exp(-\beta x)} \right)^3 \tag{29}$$

Substituting into (25) we obtain:

$$\begin{aligned} C_A^* \simeq & \frac{1}{1+e^{-\frac{\beta\sigma\sqrt{t}}{2}}} + \frac{\alpha\beta e^{-\frac{\beta\sigma\sqrt{t}}{2}}}{(1+e^{-\frac{\beta\sigma\sqrt{t}}{2}})^2} + \frac{1}{2}\alpha^2 \left[\frac{-\beta^2 e^{-\frac{\beta\sigma\sqrt{t}}{2}}}{(1+e^{-\frac{\beta\sigma\sqrt{t}}{2}})^2} + \frac{2\beta^2 e^{-\beta\sigma\sqrt{t}}}{(1+e^{-\frac{\beta\sigma\sqrt{t}}{2}})^3} \right] \\ & -d \left[\frac{1}{1+e^{\frac{\beta\sigma\sqrt{t}}{2}}} + \frac{\alpha\beta e^{\frac{\beta\sigma\sqrt{t}}{2}}}{(1+e^{\frac{\beta\sigma\sqrt{t}}{2}})^2} + \frac{1}{2}\alpha^2 \left[\frac{-\beta^2 e^{\frac{\beta\sigma\sqrt{t}}{2}}}{(1+e^{\frac{\beta\sigma\sqrt{t}}{2}})^2} + \frac{2\beta^2 e^{\beta\sigma\sqrt{t}}}{(1+e^{\frac{\beta\sigma\sqrt{t}}{2}})^3} \right] \right] \end{aligned} \tag{30}$$

Because $\beta\sigma\sqrt{t}/2$ is small in practice, we derive a linear approximation to $\exp(-n\beta\sigma\sqrt{t}/2)$ and $(1+\exp(-n\beta\sigma\sqrt{t}/2))^{-n}$, for $n = 1, 2, 3 \ldots$, by developing a Maclaurin's series for each function and ignore terms higher than degree one. We obtain[21]

$$\exp(-n\beta\sigma\sqrt{t}/2) = 1 - \frac{n\beta\sigma\sqrt{t}}{2} \quad \text{for} \quad n = 1, 2, 3 \ldots \tag{31}$$

and

$$(1+\exp(-n\beta\sigma\sqrt{t}/2))^{-n} = 2^{-n} \left(1 + \frac{n\beta\sigma\sqrt{t}}{4} \right) \quad \text{for} \quad n = 1, 2, 3 \ldots \tag{32}$$

It is important to note that $C_A^* \equiv C_A^*(\alpha, \sigma\sqrt{t})$ is a bivariate function of α and $\sigma\sqrt{t}$, both components being functions of $\sigma\sqrt{t}$, and usually with $\sigma^2 t \ll 1$. Accordingly, for fixed d, C_A^* will behave more like $1/\sigma\sqrt{t}$ than $\sigma\sqrt{t}$. As a consequence, we refine our approximation in α and expand to degree 2, but only undertake a linear expansion relative to $\sigma\sqrt{t}$. It is convenient to represent the term of the expansion $\alpha^n(\sigma\sqrt{t})^m$ by the symbol (n, m).

To derive the zero, first and second order approximations[22] for C_A^*, as well as the equations for the implied volatility, σ_i, that is $(0, 1), (1, 1)$ and $(2, 1)$, we substitute approximations (31) and (32) into (30) to obtain directly (33), (35) and (37). We replace C_A^* by its corresponding market value, C^*, and σ by σ_i, to obtain the associated implied volatility equations:
Zero order-$(0, 1)$:

$$C_A^* = \frac{1}{2}(1-d) + \frac{\beta\sigma\sqrt{t}}{8}(1+d), \tag{33}$$

[21]We note that our linear approximation always underestimates the exponential function. Further, we use the sign = to emphasis that we shall put $1 - \mathcal{O}\left((\beta\sigma\sqrt{t})^2)\right) = 1$ in all subsequent approximation analysis.

[22]It should be noted that each one of our approximations for C_A^* can be negative. Therefore, care should taken if they are used for price approximation. For options not out of-the-money, the zero and the first order approximations are always positive.

$$\sigma_i\sqrt{t} = \frac{8\left(C^* - \frac{1}{2}(1-d)\right)}{\beta(1+d)}. \tag{34}$$

First order-$(1,1)$:

$$C_A^* = \frac{1}{2}\left(1 + \frac{\alpha\beta}{2}\right)(1-d) + \frac{\beta\sigma\sqrt{t}}{8}(1+d), \tag{35}$$

$$\left(\sigma_i\sqrt{t}\right)^2 - \frac{8\left(C^* - \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) - 2\log(d)\frac{(1-d)}{(1+d)} = 0. \tag{36}$$

Second order-$(2,1)$:

$$C_A^* = \frac{1}{2}\left(1 + \frac{\alpha\beta}{2}\right)(1-d) + \frac{\beta\sigma\sqrt{t}}{8}\left(1 - \frac{\alpha^2\beta^2}{4}\right)(1+d), \tag{37}$$

$$\left(\sigma_i\sqrt{t}\right)^2 - \frac{8\left(C^* - \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) - 2\log(d)\,\frac{(1-d)}{(1+d)} - \frac{(\beta\log(d))^2}{4} = 0. \tag{38}$$

We see that the use of the logistic function to obtain $N_A(\cdot)$ results in a polynomial in α of all degrees. Because α is a linear function of the moneyness ratio d, this means that we are able to better approximate the implied parameters. In distinct contrast, if we base our approximation on (18) to derive $N_A(\cdot)$ and its derivatives (or equivalently $N(\cdot)$), we obtain an odd order power series in $\alpha \pm \sigma\sqrt{t}/2$, which prevents the derivation of simple formulae. This is because if the call approximation is developed in terms of the argument of $N(x)$, it will be an odd order power series of the same type as (15) and (18). On the other hand, noting that $N(x)$ can be viewed as a new variable that is obtained by rescaling x, we can expand the call formula in terms of $N(x)$ and obtain an expansion that will contain both odd and even terms. We shall see in the next section the importance of our methodology. It produces new and better approximations, as well as including all the standard results. Moreover, because it is very intuitive, it allows us to have good control and understanding of the approximations made, which will lead us to further major improvements.

B Results

In this sub-section we compare and contrast our results with earlier ones, and also derive our prime results. We first consider how well our approach allows us to approximate the BSM price. We then discuss the associated implied volatility estimation properties. Finally, we derive our recommended formulae, obtaining them by building on those developed directly from the logistic function approximation.

B.1 Call Price Approximation

We commence by considering the at-the-money approximations. Direct substitution of $d = 1$ into any of the equations (33),(35), or (37) recovers the result $C_A^* = \sigma\sqrt{t}/(2\pi)$ due to Brenner and Subrahmanyam (1988). This reflects the interchangeability of the logistic function and the cndf about $d = 1$ when calculating the call option value. An important point to note is that neither the zero nor the second order approximations can be derived by means of Laplace's (1785) series, (15), whereas that of the first order is equivalent to undertaking a linear approximation to $N(x)$, as used by Bharadia et al. (1996), and Corrado and Miller (1996).

Turning now to the not-at-the-money position, in Figures 1 and 2, for $t = 1/4$ and $t = 1/12$, we plot our zero, first, and second order approximations to the BSM price, C_A. Except for a very narrow region around $d = 1$, where there is a marginal improvement, the first order approximation (previously obtained by Bharadia et al. (1996), and Corrado and Miller (1996)) is worse than our zero order one. From a computational view, our zero order approximation is far simpler than either of the other two approximations. However, our second order approximation does introduce a very significant improvement in accuracy, and is no more computationally intensive than that of the inferior first order one.

B.2 Call Implied Volatility Approximation

We now focus our attention on the calculation of the implied volatility, σ_i, but first derive a general quadratic equation that contains (34),(36), and (38)as special cases. It is elementary that for $d > 1$, $\log(d) \simeq 2\left(\frac{d-1}{d+1}\right)$, if powers of $(d-1)/(d+1)$ greater than or equal to three are ignored. Defining $\delta = 0$, 4 and $(4 - \frac{8}{\pi})$, we can write (34),(36), and (38) in the compact form:

$$\left(\sigma_i \sqrt{t}\right)^2 - \frac{8\left(C^* - \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) + \delta\left(\frac{1-d}{1+d}\right)^2 = 0. \tag{39}$$

We note that the zero order approximation equation for σ_i is linear in $\sigma_i\sqrt{t}$, in contrast with the first and second order approximations. We further define[23]:

$$b \equiv \frac{4\left(C^* - \frac{1}{2}(1-d)\right)}{\beta(1+d)} = \frac{2}{\beta}\left(\frac{C+P}{S\exp(-at) + X\exp(-rt)}\right), \tag{40}$$

and

$$m \equiv \left(\frac{1-d}{1+d}\right)^2. \tag{41}$$

By writing (40) as $(\sigma_i\sqrt{t})^2 - 2b\sigma_i\sqrt{t} - \delta m = 0$, we obtain the following *simple* formula for the implied volatility, σ_i

$$\sigma_i = \frac{b \pm \sqrt{b^2 - \delta m}}{\sqrt{t}} \tag{42}$$

where σ_i has two possible roots. To find the unique solution we first note that any solution must be valid for all d about $d = 1$. For $d = 1$ (42) becomes

[23] In the Appendix we discuss the economic meaning of the "market prices" representation of b.

$$\sigma_i = \frac{b \pm b}{\sqrt{t}} = \begin{cases} \frac{2b}{\sqrt{t}} & \text{"positive" root} \\ 0 & \text{"negative" root} \end{cases} \tag{43}$$

The "negative" root is clearly not economically meaningful as it represents no uncertainty and, therefore, the "positive" root is our desired solution. This confirms the intuition of previous authors.

To test the efficiency of our results we first generated, for different values of d, "standardised" call option prices, C^*, from (4), conditioned on a given volatility, $\sigma_1 = 0.3$, and a fixed strike price, $X = 50$, (this is equivalent to varying S). Using (42) we then calculated the corresponding implied volatility for each value of C^* obtained. Our results are reported in Tables 1 and 2, and are graphically presented in Figures 3 and 4. This gives us a simple benchmark with which to compare, contrast, and check the quality of our different approximations, as σ_1 is the true volatility and is independent of S.

As expected, we see that the Brenner and Subrahmanyam (1988) approximation is good for $d = 1$, but it cannot really be used otherwise. Not surprisingly, our zero, first, and second order approximations bring a significant improvement around $d = 1$. In Figures 5 and 6 we plot σ_i against S, in the ranges $\pm 1.01\sigma_1$, and $S \in [40, 60]$ (roughly equivalent to $d \in [0.8, 1.2]$). We see that the curvatures of the zero and first order approximations are about the same, whereas the curvature of the second order one is significantly smaller[24], implying improved results. The zero and second order approximations are convex functions of S, and are valid over a large range of S, while in contrast, the first order approximation is concave in S, and is only valid in a rather narrow region of S.

At $d = 1$ all the approximation formulae reduce to that of Brenner and Subrahmanyam (1988). However, the implied volatility, in this case, necessarily underestimates the true volatility. As one moves away from $d = 1$, the zero and second order approximations, being convex in d, become increasingly more accurate. In contrast, as d moves away from unity, the first order approximation becomes rapidly more inaccurate, as it is concave[25] in d. As a consequence, we see that the first order approximation, first derived by Bharadia et al. (1996), and Corrado and Miller (1996), is the one with the worst trade off between accuracy and complexity of calculation. The zero order approximation gives slightly better results than that of the first order and only requires the solution of a linear equation. On the other hand, the second order approximation gives significantly better results, but requires the solution of a quadratic equation, as is the case for the first order approximation. Clearly at this stage we must recommend the use of either the zero or second order approximation formulae, and cannot recommend the use of the first order approximation formula. The choice of zero order against second order approximation is a trade-off between accuracy and computational complexity. Before discussing this further, we show how to improve our quadratic formula.

B.3 An Improved Implied Volatility Quadratic Approximation

We first note that increasing the order of our approximations has the effect of modifying the sign, as well as the size of the curvature of $\sigma_i \equiv \sigma_i(S)$ (viewing σ_i as a function of S). A decrease in the amplitude of the curvature gives a flatter curve and, therefore, a better

[24]It should be noted that σ_1, being independent of S, has zero curvature.

[25]It is important to keep in mind that when the best estimator underestimates, concave approximations will *always* underestimate the exact value of the volatility.

approximation. The natural question to ask is whether it is possible to use this information to improve our approximation scheme? The answer is yes, and for that we return to (39). Now formula (39) is a compact way to express σ_i for the zero, first, and second order approximations, as a function of the single parameter δ. The value of δ depends on the order of the approximation and so "controls" the curvature of σ_i. Thus our task is to find that value of δ that minimizes the curvature at $d = 1$, which is equivalent to increasing the order of our approximation, but without increasing the degree of the equation for σ_i.

Now from elementary calculus, if $y \equiv y(x)$, the curvature is defined to be $1/\rho$, where ρ is the radius of curvature given by the formula:

$$\rho = \frac{(1 + (y')^2)^{\frac{3}{2}}}{y''} \tag{44}$$

At a point of inflexion $y' = y'' = 0$ and thus ρ becomes infinite; therefore the curvature at a point of inflexion is zero. Hence, we can find the optimal δ, δ_{opt}, by imposing on $\sigma_i(S)$, at $d = 1$, the following two conditions:

$$\left.\frac{\partial\left(\sigma_i\sqrt{t}\right)}{\partial S}\right|_{d=1} = 0 \quad \text{and} \quad \left.\frac{\partial^2\left(\sigma_i\sqrt{t}\right)}{\partial S^2}\right|_{d=1} = 0. \tag{45}$$

Partially differentiating (42) with respect to S, we find that the first derivative is equal to zero, at $d = 1$, if[26]

$$\frac{\partial C}{\partial S} = \frac{1}{2} + \frac{C}{2S}. \tag{46}$$

Similarly, partially differentiating (42) twice with respect to S and evaluating at $d = 1$, the second derivative condition of (45) is satisfied only if

$$\delta_{opt} = \sqrt{8\pi}\, n\left(\frac{\sigma_i\sqrt{t}}{2}\right) \tag{47}$$

where $n(x) = \exp(-x^2/2)/\sqrt{2\pi}$. As we have assumed that $\sigma_i\sqrt{t} < 1$, we have

$$\delta_{opt} \simeq 2\left(1 - \frac{\left(\sigma_i\sqrt{t}\right)^2}{8}\right). \tag{48}$$

Substituting (48) into (39), and collecting terms, we obtain a new quadratic formula for the implied volatility

$$\sigma_i = \frac{\frac{b}{\left(1-\frac{m}{4}\right)} + \sqrt{\left(\frac{b}{\left(1-\frac{m}{4}\right)}\right)^2 - \frac{2m}{\left(1-\frac{m}{4}\right)}}}{\sqrt{t}}. \tag{49}$$

[26]In fact, this is the condition that b has a turning point at $d = 1$. From (40) we have that $b = 2/\beta(C + P)/(S\exp(-at) + X\exp(-rt))$. Differentiating b partially with respect to S, and applying the Put-Call Parity Theorem gives (46), when $d = 1$. It should be noted that (46) is explicitly independent of δ and that is because $\partial m/\partial d = 0$ when $d = 1$.

If $b^2 - 2m(1 - m/4) \geq 0$ we can expand (49) via the Binomial Theorem. If b^2 is sufficiently large, and we can also ignore $(m/4)^2$ and higher order terms, we are able to undertake a linear approximation to obtain the *simple* formula:

$$\sigma_i \sqrt{t} = b(2 + m/2) - m/b \; . \tag{50}$$

In Figures 5 and 6 we plot (49) and (50). It is clear that the approximation is greatly improved by using δ_{opt} in (39). Also, although very simple, (50) gives very good results. In the next sub-section we discuss these results in more detail.

The "control" method used is equivalent to summing all the higher order terms and incorporating the information in a redefined δ. This is a very powerful approach, which is very close to what physicists call a renormalization technique. A similar method is used by Corrado and Miller (1996). Although their linear approximation is significantly improved by the method, some aspects of their derivation must be taken with care. They appear to assume that the implied volatility must be a linear function of S, whereas it must be independent of S. Moreover, in their derivation of what we call an optimum δ, they make an error[27]. They then check how their δ_{opt} varies as a function of $\sigma_i \sqrt{t}$ and choose one *convenient* number in the range of variation, arguing that this range is small for practical values of $\sigma_i \sqrt{t}$. This may be true, but the range they obtain is smaller than it should be, due to their error in calculating the value of σ_i, which is also rather sensitive to variations in δ_{opt}. It is therefore better to take this dependency into account, as in (48), especially as it improves the results without introducing further computational intensity. Fortunately, the value chosen by Corrado and Miller (1996), $\delta_{opt} = 2$, "for simplicity", corresponds in fact to the zero order approximation of (48), which allows them to significantly improve their results.

We can even go one step further. The $\sigma_i(S)$ we obtained from (49) is a concave function of S and therefore, as previously mentioned, will underestimate the true volatility for all values of S. We are able to counter some of the effect introduced by this bias in our approximation by modifying the factor 2 in the last term of expression (49). By requiring that σ_i should be within $\pm 1\%$ of the true value, we found that a factor of 1.875 gives extremely good results, as we can see from Figures 3, 4, 5 and 6. For options with maturities of about one month we have an error of less than 1% for $d = 1 \pm 0.1$, whilst for three months maturities the same error condition gives $d = 1 \pm 0.2$. Following the same argument, (50) is also improved, as can be seen in Figures 5 and 6, as well as Tables 1 and 2.

C Discussion of the Formulae and Recommendations

In this sub-section[28] we now discuss Tables 1 and 2 in detail, and put forward our recommendations for the implementation of our formulae. In the first column of each table we give the implied volatility based on the Brenner and Subrahmanyam formula over the underlying's range $[40, 60]$, when the strike price = 50 and the true volatility is 30%. In this case, $d = 1$ corresponds to $S = 49.26$ when $t = 1/4$, and $S = 49.75$ when $t = 1/12$. We see that the Brenner and Subrahmanyam performance deteriorates remarkably rapidly. We cannot

[27]Expressions (8) and (9) in Corrado and Miller (1996) consist of two terms, the second of which, in their notation is $\pi[\phi(\sigma\sqrt{T}/2 - 1/2]^2$, should not appear. In their reported results they appear to use the incorrect expressions.

[28]Our analysis assumes that $a = 0$ without loss of generality.

envisage the circumstances in which one would want to use this approximation in practice. In the following three columns we give the implied volatility results obtained by using our order approximations, (34), (36), and (38), respectively. All three are obviously superior to that of Brenner and Subrahmanyam. It is interesting to note that the zero order formula dominates the first order approximation whenever the first order formula can be used[29].

The column headed "HS opt2" gives the implied volatility obtained from the "optimised" version of equation (49). We explained in the previous sub-section's ultimate paragraph that if $1.875m$ replaces the term $2m$, then we obtain our best approximation. For exchange traded equity options, for example, the increments between strikes is about ten per cent of the in- and out-of-the-money strike values. Thus in our case, when the strike level is about 500, we look at the implied volatility behaviour up to at least 550, and down to at least 450. As we can see, to two decimal places we cover the whole strike price range[30] $[450, 550]$ and, in fact, go further. Clearly, if one includes our second order approximation in relevant software, this is the formula that we should use. That is:

$$\sigma_i = \frac{\frac{b}{\left(1-\frac{m}{4}\right)} + \sqrt{\left(\frac{b}{\left(1-\frac{m}{4}\right)}\right)^2 - \frac{1.875m}{\left(1-\frac{m}{4}\right)}}}{\sqrt{t}}. \tag{51}$$

Although (51) is a simple expression, and would certainly be used as part of a computer code, we derived a less complicated one above, expression (50). We now ask how good is (50) and how good is its "optimised" version? These questions are addressed by the information contained in the columns headed "HS(C&P)" and "HS(C&P)opt", respectively. Once again, using the two decimal place criterion, we see that both approximations are excellent up to at least half a strike price increment, either side of the at-the-money position. The accuracy behaviour of "HS(C&P)" for the short dated option, Table 2, reflects the option's greater convexity relative to that of the longer dated one, Table 1, where we see that "HS(C&P)opt" is as accurate as the more complicated expression "HS opt2". We conclude, therefore, that traders in open out cry markets, for example, can confidently use (50), and this is what we recommend.

D European Put Option Implied Volatility Estimation

The focus of this paper is on European style put and call options. They have similar characteristics, but do have some important differences. For example, puts have limited upside potential in contrast with calls that have unlimited profitability. Further, calls have an elasticity, with respect to the underlying, which is never less than one, whilst the absolute value of a put's equivalent elasticity can be less than unity. Consequently, we believe that it is reasonable to expect that the functional form of a put's implied volatility estimator will necessarily differ from that of a call option's written on identical terms. We shall now show that this is not the case. In fact, we shall prove the key result that the functional form of our implied volatility estimator is the same irrespective as to whether a put or call option is being

[29]It should be noted that the first order approximation is valid only over a rather limited range, as the implied volatility must be real. Bharadia et al. (1996), and Corrado and Miller (1996) essentially independently derived this form of approximation.

[30]Indeed, to three decimal places, we are never more than 0.34% in error. In practice, one would almost certainly never work to more than two decimal places.

analyzed. However, we shall not give the detailed calculations, because they are essentially the same as those presented in the previous section on the call option. We do present though, for the readers convenience, the formulae when the input data are based on put options.

E The Put Option Approximation Formulae

From (2), the BSM put option valuation formula, we proceed as for the call option, and by replacing P_A^* by its corresponding market value, P^*, and σ by σ_i, where appropriate, we obtain the zero, first and second order approximations for P_A^* and the implied volatility, σ_i:

Zero order-$(0,1)$:

$$P_A^* = -\frac{1}{2}(1-d) + \frac{\beta\sigma\sqrt{t}}{8}(1+d), \tag{52}$$

$$\sigma_i\sqrt{t} = \frac{8\left(P^* + \frac{1}{2}(1-d)\right)}{\beta(1+d)}. \tag{53}$$

First order-$(1,1)$:

$$P_A^* = \frac{1}{2}\left(-1+\frac{\alpha\beta}{2}\right)(1-d) + \frac{\beta\sigma\sqrt{t}}{8}(1+d), \tag{54}$$

$$\left(\sigma_i\sqrt{t}\right)^2 - \frac{8\left(P^* + \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) - 2\log(d)\frac{(1-d)}{(1+d)} = 0. \tag{55}$$

Second order-$(2,1)$:

$$P_A^* = \frac{1}{2}\left(-1+\frac{\alpha\beta}{2}\right)(1-d) + \frac{\beta\sigma\sqrt{t}}{8}\left(1-\frac{\alpha^2\beta^2}{4}\right)(1+d), \tag{56}$$

$$\left(\sigma_i\sqrt{t}\right)^2 - \frac{8\left(P^* + \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) - 2\log(d)\,\frac{(1-d)}{(1+d)} - \frac{(\beta\log(d))^2}{4} = 0. \tag{57}$$

It is more than interesting to note that the approximations for P_A^*, can obtained by substituting for C_A^* through the put-call parity relationship (8) and applying it to (33-38). Therefore, the results for the put option can be viewed as though we had made a direct application of the Put-Call Parity Theorem to derive the explicit put approximations (which was not the case!).

There is essentially an anti-symmetric relationship between (33-38) and (52-57) induced by replacing C by P. For example, in the second order approximation the first term of (56) has -1 instead of $+1$, and in (57) $-1/2(1-d)$ is replaced by $+1/2(1-d)$. The various price approximation results are very similar to those for the call option and are shown in Figure 7. Accordingly, in parallel with the analysis leading to (39), we can now rewrite (52-57) in the compact form:

$$\left(\sigma_i\sqrt{t}\right)^2 - \frac{8\left(P^* + \frac{1}{2}(1-d)\right)}{\beta(1+d)}\left(\sigma_i\sqrt{t}\right) + \delta\left(\frac{d-1}{d+1}\right)^2 = 0, \tag{58}$$

where, as before, $\delta = 0$, 4 and $(4 - \frac{8}{\pi})$ give the zero, first and second order approximations respectively. Using the Put-Call Parity Theorem (8) we see that

$$\frac{4\left(P^* + \frac{1}{2}(1-d)\right)}{\beta(1+d)} = \frac{2}{\beta}\left(\frac{C+P}{S\exp(-at) + X\exp(-rt)}\right) = b \tag{59}$$

Substituting (59) into (58) shows that (58) is in fact equal to (39). Thus our implied volatility put formulae are identical to those that we obtained for the call. b can be written more compactly as

$$b = \frac{2}{\beta}\left(\frac{C^* + P^*}{1+d}\right) \tag{60}$$

As a result, we clearly see that we need to value a straddle for a given strike price[31], whilst the denominator reflects the moneyness of the option[32].

With hindsight, we should not be totally surprised by these results. There are two key points to bear in mind. First, in Section III our Theorem suggests that the underlying symmetry of the BSM formula and of the logistic function is likely to be transmitted to the associated estimators. Second, and possibly more importantly, b is a symmetric function of $S\exp(-at)$ and $X\exp(-rt)$ jointly, with the numerators and denominators also being symmetric functions in their own right. We see that from (40) and (59) the r.h.s. of (40) is the natural representation of b, rather than that of the more compact form (60), which is obtained directly from our quadratic equations for the implied volatility[33]. However, b is not a symmetric function of $S\exp(-at)$ or $X\exp(-rt)$ individually, and this fact is vital in understanding where the put or call properties are in our formulae.

Now for fixed $X\exp(-rt)$, b has a unique minimum when $S\exp(-at) = X\exp(-rt)$, that is, when $d = 1$, and is asymptotic[34] to $\sqrt{\pi/2}$ as $S\exp(-at) \to 0$ or ∞. As a result, b is a convex function around $d = 1$, but has points of inflexion such that as $S\exp(-at)$ continues to increase or decrease, b now becomes a concave function. It is obvious that $S\exp(-at)$ must tend to zero more rapidly than to infinity and, therefore, b must be asymmetric with respect to $S\exp(-at)$. Because of the joint symmetry in $S\exp(-at)$ and $X\exp(-rt)$ this means that b must also be asymmetric with respect to $X\exp(-rt)$. When $S\exp(-at) \to 0$ we clearly must have $S\exp(-at) < X\exp(-rt)$ and therefore, the value of the put is greater

---

[31]There will usually be two different strikes that will give the same straddle value.

[32]Loosely, $1 + d$ can be viewed as a deflator (inflator) function, depending on whether $d - 1 > 0(< 0)$. However, great care must be taken with this interpretation as b is not a monotonic function of d.

[33]This type of result is not new. In Garman (1983) it is shown that the *... option pricing through forward prices, is somehow more fundamental than the original option pricing approach (Black and Scholes (1973)) ...*'

[34]Although we use the word asymptotic and associated symbolism, this is merely a convenience. We do not imply that our approximation formulae are valid over the range $[0;\infty]$. Our comments relate only to the mathematical properties of b.

than that of the call. Hence, not surprisingly, when the put is in-the-money its influence must dominate that of the call, but the important point is that this influence is not symmetric about $S\exp(-at) = X\exp(-rt)$. As a result, although the implied volatility formula is the same for both puts and calls, and is jointly symmetric in $S\exp(-at)$ and $X\exp(-rt)$, the fundamental asymmetry of the options' payoffs comes into play through the asymmetric property of b.

V Summary and Conclusion

The focus of this paper has been on the derivation of computationally simple formulae for the determination of the implied volatility of a plain vanilla European put or call option, when a continuous dividend is paid directly proportional to the underlying, and when the options are not at-the-money. The basic principle that we have used is approximation by surrogates. Our methodology can be easily adapted to determine all implied parameters, when needed either singly or jointly. We have shown that by the use of what is effectively an exponential sum rational approximation, in this case specialised to the logistic function, we are able to produce formulae with low computational intensity that are simple in form and easy to use. They are never inferior with respect to previous formulae, from both a speed and accuracy view. Given these two criteria, our formulae dominate those presented previously. Further, our formulae are the same irrespective as to whether the required implied volatility is that of a put or a call. We have explained this result mathematically and given an economic explanation, as well as also resolving various issues raised by previous research. Finally, the validity regions for our approximation formulae are such that we believe that our results will be of great benefit to traders, as well as managers of portfolios containing a large number of derivatives, and shed further light on the issues raised by Bodie (1995).

Appendix: The Economic Interpretation of b

A fundamental property of plain vanilla put and call options is that their payoffs can be replicated by simple combinations of long or short positions in the underlying asset, and access to riskless (in the sense of default) borrowing and lending. In particular, a long call option payoff is replicated by a long position in the underlying and a short position in a riskless zero-coupon bond, with face value equal to the option's strike price. For a similar put option, the converse portfolio is a long position in the bond and a short position in the underlying. We see, therefore, that a two-asset portfolio consisting of an investment in a single risky asset and a risk-free asset can be viewed as a *fundamental* portfolio[35] in derivatives analysis. Indeed, this portfolio, with positive investments in each asset, may be viewed as a *composite numéraire*, as it never takes the value zero, and acts as a reference base for other investment strategies.

A key aspect of the payoff characteristic of our numéraire portfolio is that its payoff is linear with respect to the underlying. This is in distinct contrast with an option's payoff that, by construction, is non-linear. If we construct a portfolio that has a non-linear payoff, which we shall term a *derivatives portfolio*, and wish to determine the effect of its skewness and convexity properties, for example, it seems natural to compare it with our numéraire portfolio, which has neither of these properties. Comparing the ratio of their market values[36], we obtain a dimensionless variable, which must reflect the fundamental financial economics differences between the numerator and the denominator. When the prime difference between the derivatives portfolio and the numéraire portfolio is the convexity created by the (embedded) options, the source of that convexity is the underlying's volatility. Hence the ratio must directly reflect the implied volatility.

In this Appendix we consider the function

$$B(S\exp(-at), X\exp(-rt)) \equiv \frac{C+P}{S\exp(-at) + X\exp(-rt)} \equiv \frac{\beta}{2}\, b \tag{61}$$

that arises from the fundamental implied volatility call equation (39)

$$\left(\sigma_i\sqrt{t}\right)^2 - 2\, b\left(\sigma_i\sqrt{t}\right) + \delta\left(\frac{d-1}{d+1}\right)^2 = 0 \tag{62}$$

where[37]

$$b \equiv \frac{4}{\beta}\;\frac{\left(C^* - \frac{1}{2}(1-d)\right)}{(1+d)} \tag{63}$$

and drives all of our implied volatility estimators. It should be noted that all approximations, based on logistic function expansion about $d = 1$, include $B(\cdot,\cdot)$. In fact, the higher order approximations essentially reflect the need to add appropriate "correction" terms to $B(\cdot,\cdot)$.

[35]Since the development of the CAPM one of the principle areas of research has been that of mutual fund separation, especially two-fund separation, where one of the funds represents a riskless, in the sense of default, bond and the other the Market Portfolio.

[36]This ratio can be viewed as a *return on capital employed*

[37]We remind the reader that we proved in Subsection IV.D that the equation for the implied volatility is the same for a European call or put option.

$B(\cdot, \cdot)$ is a symmetric function of $S\exp(-at)$ and $X\exp(-rt)$, that is,

$$B(S\exp(-at), X\exp(-rt)) = B(X\exp(-rt), S\exp(-at)). \tag{64}$$

Further, both the numerator and denominator are symmetric functions in their own right. The numerator, $C+P$, is a *straddle*. This is the classic volatility trade used by options traders and has, of course, convexity properties. The denominator, however, has no convexity, and can be viewed as the fundamental two-asset portfolio at the heart of Modern Portfolio Theory. The wealth allocation decision directly reflects the straddle's strike price, which must be first obtained by determining the optimal (or given) strike price for the straddle. This contrasts with the usual wealth allocation decision that is determined directly from the individual agent's utility function. From the above analysis it is clear that $B(S\exp(-at), X\exp(-rt))$ must directly reflect the implied volatility of the option.

References

Bell, E. T. *Men of Mathematics*, Vol. 2, Reprinted 1965, Pelican Books, England (1937).

Bharadia, M. A. J., Nicos Christofides, and Gerry Salkin. "A Quadratic Method for the Calculation of Implied Volatilty using the Garman-Kohlhagen Model." *Financial Analysts Journal*, (March-April) (1996), 61–64.

Black, Fischer and John C. Cox. "Valuing Corporate Securities: Some Effects of Bond Indenture Provisions." *Journal of Finance*, 31 (May) (1976), 351–368.

Black, Fischer and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy*, 81 (May-June) (1973), 637–654.

Bodie, Zvi."On the Risk Stocks in the Long Run." *Financial Analysts Journal*, (May-June) (1985), 18–22.

Brenner, Menachem and Marti G. Subrahmanyam. " A Simple Formula to Compute the Implied Standard Deviation." *Financial Analysts Journal*, (September-October) (1988), 80–83.

Brenner, Menachem and Marti. G. Subrahmanyam. "A Simple Approach to Option Valuation and Hedging in the Black-Scholes Model." *Financial Analysts Journal*, (March-April) (1994), 25–28.

Chance, Don M."A Generalized Simple Formula to Compute the Implied Volatility." *The Financial Review*, 13 (November) (1996), 859–867.

Corrado, Charles J. and Thomas W. Miller Jr. "A Note on a Simple, Accurate Formula to Compute Implied Standard Deviations." *Journal of Banking and Finance*, 20 (1996), 595–603.

Finney, David J. *"Statistical Method in Biological Assay"*, 3rd Edition, Charles Griffin & Company Ltd., London and High Wycombe, England (1978).

Garman, Mark B. "Forward Prices, Option Prices, and Dividend Corrections." unpublished working paper (1983).

Johnson, Herb. "Options on the Maximum or the Minimum of Several Assets." *Journal of Financial and Quantitative Analysis*, 22 (September) (1987), 277–283.

Johnson Norman L., Samuel Kotz, and N. Balakrishnan. "*Continuous Univariate Distributions*", (Vol. 1), John Wiley & Sons, Inc., New York (1994).

Johnson Norman L., Samuel Kotz, and N. Balakrishnan. "*Continuous Univariate Distributions*", (Vol. 2), John Wiley & Sons, Inc., New York (1995).

Laplace, Pierre S. "Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres." *Histoire de l'Académie Royale des Sciences de Paris* (1785), 1–88.

Latané, Henry A. and Richard J. Rendleman, Jr. "Standard Deviations of Stock Price Ratios Implied in Option Prices." *Journal of Finance*, 31 (May) (1976), 369–382.

Margrabe, William. "The Value of an Option to Exchange One Asset for Another." *Journal of Finance*, 33 (March) (1978), 177–186.

Merton, Robert C. "Theory of Rational Option Pricing." *Bell Journal of Economics and Management Science*, 4 (1973), 141–183.

Merton, Robert C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance*, 29 (May) (1974), 449–470.

Merton, Robert C. *Continuous-Time Finance*, Blackwell Publishers, Cambridge, Mass., U.S.A (1990).

Samuelson, Paul A. "Rational Theory of Warrant Pricing." *Industrial Management Review*, 6 (Spring) (1965), 13–31.

Stephan, Jens A. and Robert E. Whaley. "Intraday Price Changes and Trading Volume Relations in the Stock and Stock Options Markets." *Journal of Finance*, 45 (1990), 191–220.

Snelling, Paula L. "Earnings Surprises and the Information Content of Option Prices." Unpublished PhD dissertation, University of North Carolina at Chapel Hill (1987).

Varson, Paula L. and Michael J.P. Selby. "Option Prices as Predictors of Stock Prices: Intraday Adjustments to Information Releases." *The European Journal of Finance*, 3 (1997), 49-72.

Table 1

Value of the Implied Volatility for the Different Approximations

$K = 50 \quad \sigma = 0.300 \quad r = 0.06 \quad t = 1/4$							
S	BS σ	HS(0) σ	HS(1) σ	HS(2) σ	HS opt2 σ	HS(C&P) σ	HS(C&P)opt σ
40	0.032	0.547	-	0.387	-	0.392	0.367
41	0.044	0.498	-	0.366	-	0.365	0.345
42	0.060	0.453	-	0.348	0.287	0.343	0.326
43	0.080	0.414	-	0.334	0.297	0.326	0.313
44	0.104	0.380	-	0.323	0.300	0.314	0.304
45	0.132	0.352	-	0.315	0.301	0.306	0.299
46	0.165	0.330	0.258	0.308	0.301	0.302	0.298
47	0.202	0.314	0.283	0.304	0.300	0.300	0.298
48	0.243	0.304	0.295	0.301	0.300	0.300	0.299
49	0.288	0.300	0.300	0.300	0.300	0.300	0.300
50	0.337	0.301	0.298	0.300	0.300	0.300	0.299
51	0.389	0.308	0.291	0.302	0.300	0.300	0.299
52	0.444	0.319	0.277	0.305	0.301	0.301	0.298
53	0.501	0.335	0.248	0.310	0.301	0.303	0.298
54	0.560	0.354	-	0.315	0.301	0.307	0.300
55	0.620	0.377	-	0.322	0.300	0.313	0.303
56	0.681	0.403	-	0.331	0.298	0.322	0.310
57	0.743	0.431	-	0.340	0.294	0.333	0.318
58	0.805	0.461	-	0.351	0.283	0.347	0.329
59	0.867	0.493	-	0.364	-	0.363	0.343
60	0.928	0.526	-	0.377	-	0.380	0.358

Table 1: BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50); HS(C&P)opt σ: simple optimized approximation

Table 2

Value of the Implied Volatility for the Different Approximations

$K = 50 \quad \sigma = 0.300 \quad r = 0.06 \quad t = 1/12$							
S	BS σ	HS(0) σ	HS(1) σ	HS(2) σ	HS opt2 σ	HS(C&P) σ	HS(C&P)opt σ
40	0.002	0.944	-	0.605	-	0.647	0.602
41	0.004	0.840	-	0.542	-	0.576	0.537
42	0.008	0.740	-	0.483	-	0.510	0.475
43	0.015	0.645	-	0.433	-	0.450	0.420
44	0.028	0.558	-	0.391	-	0.398	0.373
45	0.048	0.480	-	0.359	0.265	0.356	0.337
46	0.078	0.414	-	0.335	0.297	0.326	0.313
47	0.119	0.362	-	0.318	0.301	0.309	0.301
48	0.173	0.325	0.268	0.307	0.301	0.302	0.298
49	0.241	0.304	0.295	0.301	0.300	0.300	0.299
50	0.322	0.300	0.299	0.300	0.300	0.300	0.300
51	0.416	0.312	0.286	0.303	0.301	0.300	0.299
52	0.520	0.338	0.241	0.311	0.301	0.304	0.298
53	0.632	0.377	-	0.323	0.301	0.313	0.303
54	0.751	0.425	-	0.339	0.296	0.330	0.316
55	0.874	0.482	-	0.359	0.265	0.356	0.337
56	0.998	0.543	-	0.385	-	0.389	0.366
57	1.123	0.609	-	0.415	-	0.428	0.400
58	1.247	0.677	-	0.449	-	0.470	0.438
59	1.369	0.746	-	0.487	-	0.514	0.479
60	1.488	0.816	-	0.527	-	0.560	0.522

Table 1: BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50); HS(C&P)opt σ: simple optimized approximation

Figure 1: Value of a call option $C(S)$ for $t = 1/4$. Black-Scholes is given by (1), HS by (24)$\cdot$ S, HS(0) by (33)$\cdot$ S, HS(1) by (35)$\cdot$ S and HS(2) by (37)$\cdot$ S.

Figure 2: Value of a call option $C(S)$ for $t = 1/12$. Black-Scholes is given by (1), HS by (24)$\cdot$ S, HS(0) by (33)$\cdot$ S, HS(1) by (35)$\cdot$ S and HS(2) by (37)$\cdot$ S.

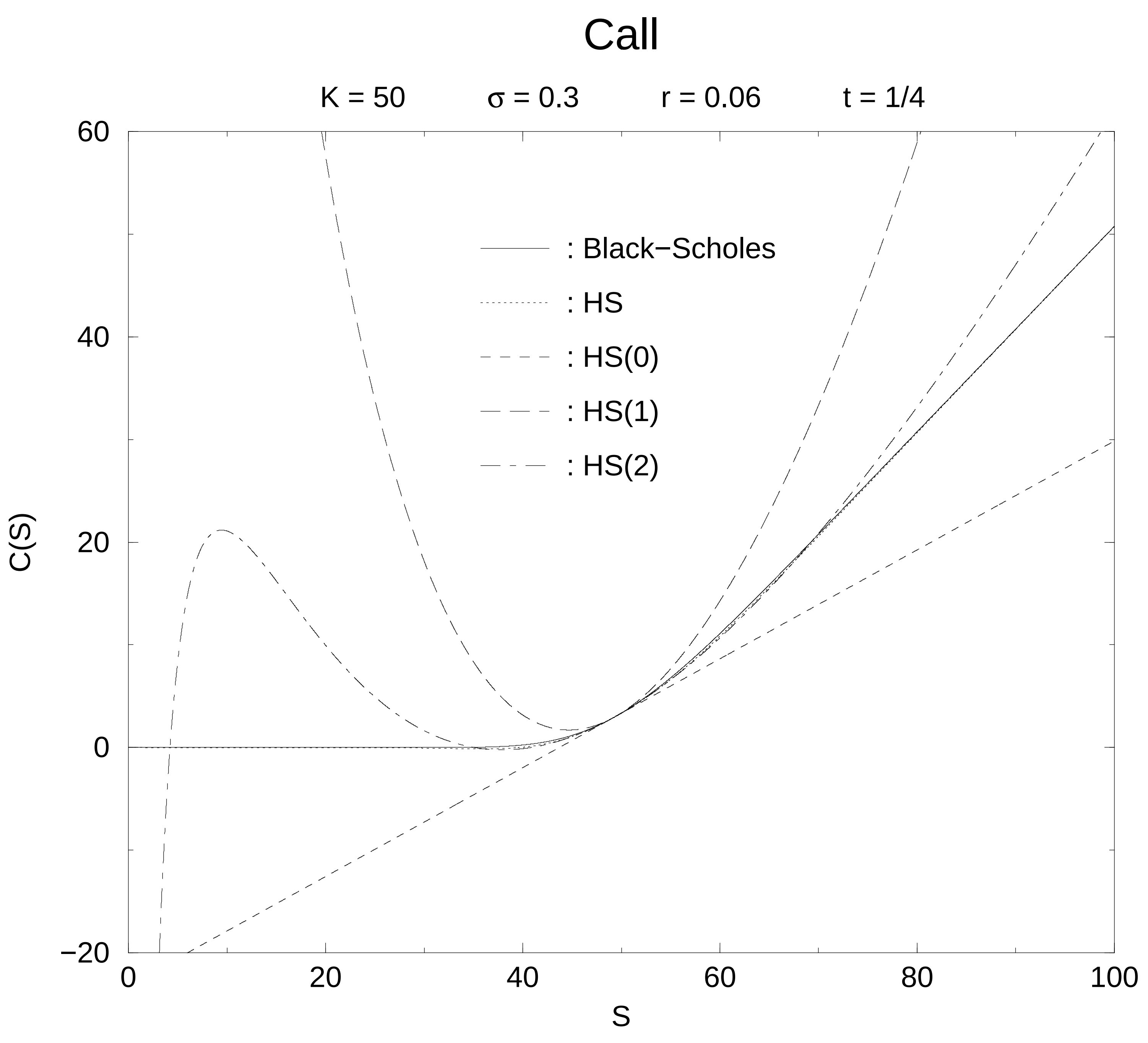
Figure 3: Implied volatility σ_i of a call option for $t = 1/4$. BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50).

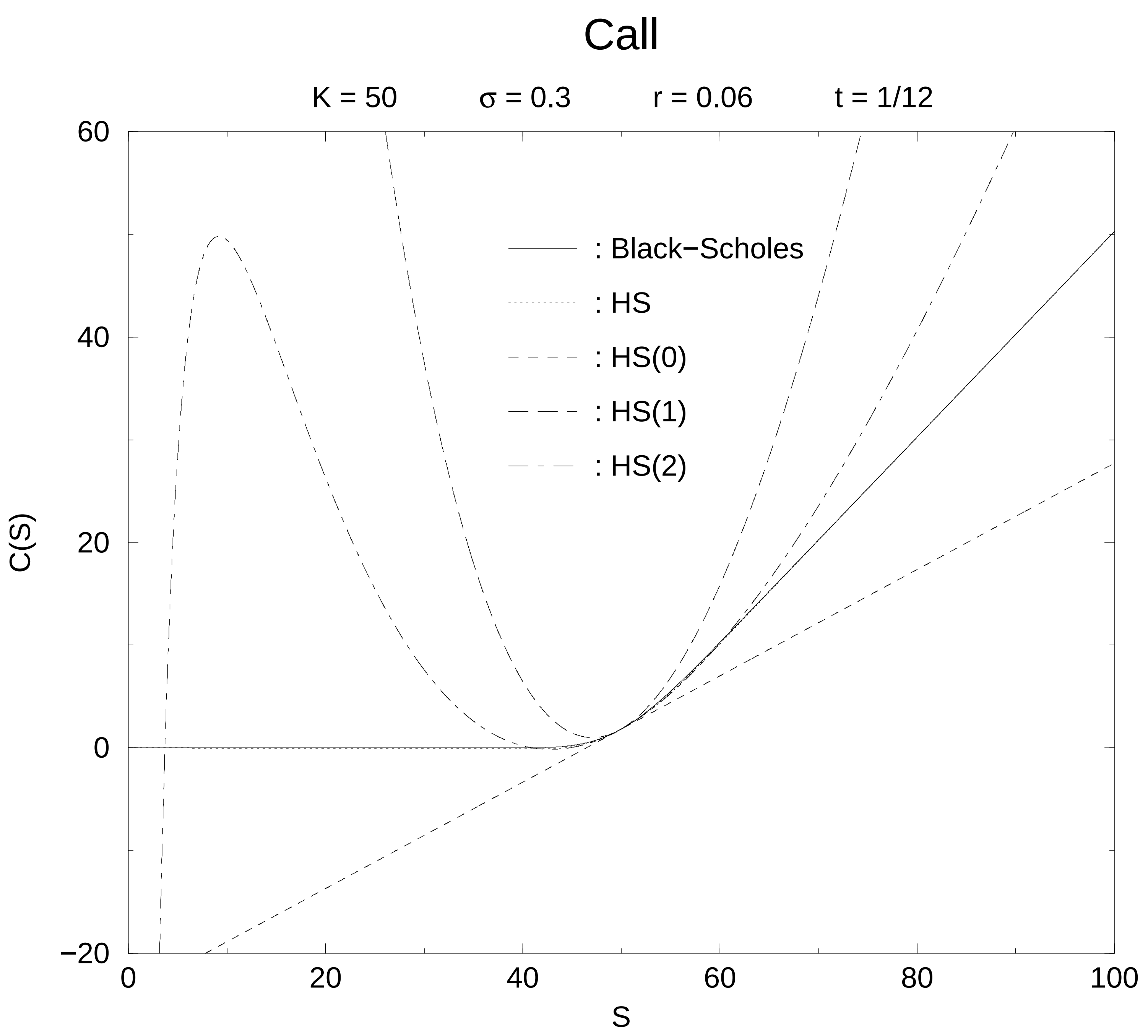
Figure 4: Implied volatility σ_i of a call option for $t = 1/12$. BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50).

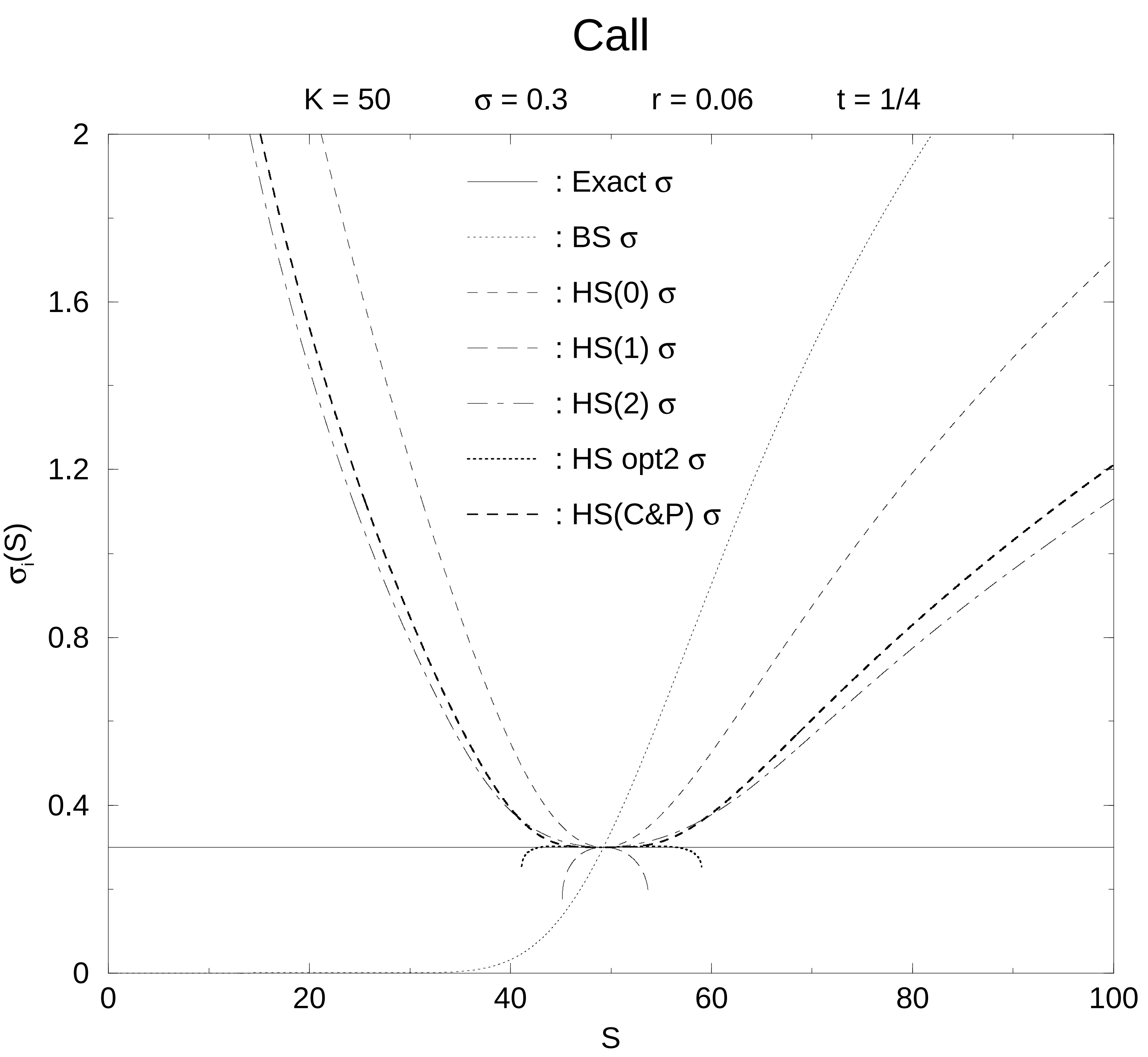
Figure 5: Magnification of the implied volatility σ_i of a call option for $t = 1/4$. BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50).

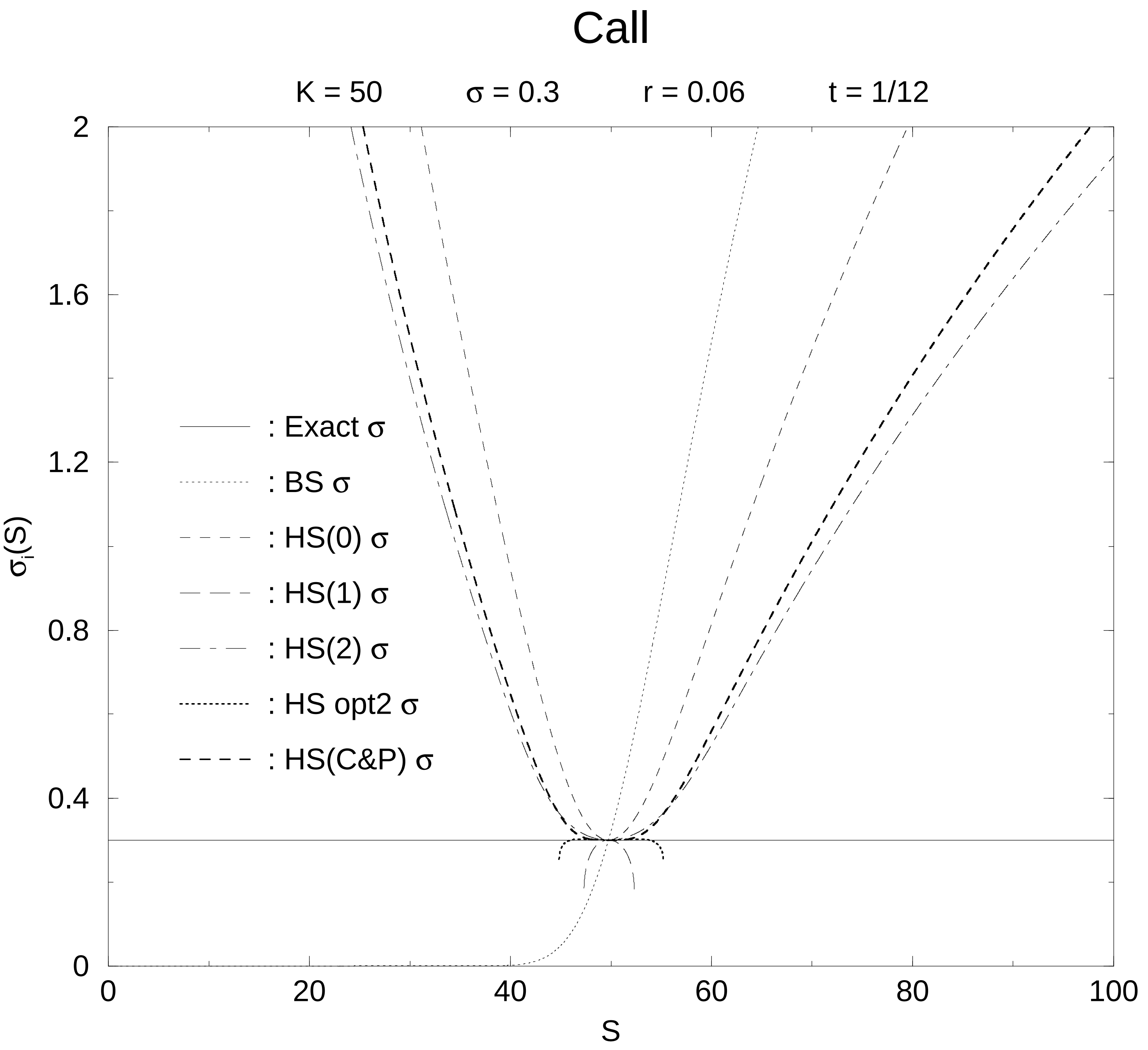
Figure 6: Magnification of the implied volatility σ_i of a call option for $t = 1/12$. BS σ: Brenner and Subrahmanyam (1988) approximation; HS(0) σ: zero order approximation (34); HS(1) σ: first order approximation (36); HS(2) σ: second order approximation (38); HS opt2 σ: optimized approximation (51); HS(C&P) σ: simple approximation (50).

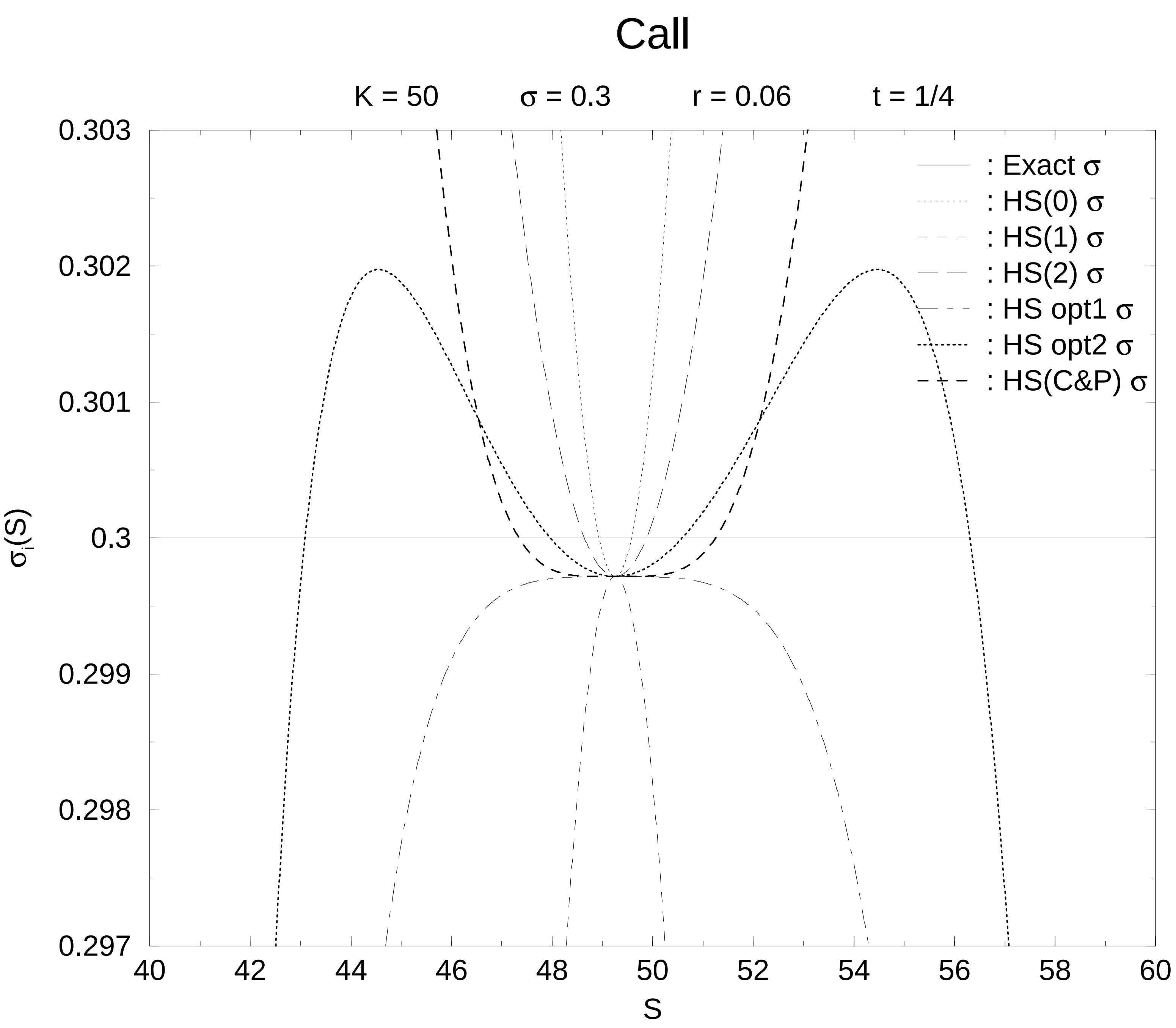
Figure 7: Value of a put option $P(S)$ for $t = 1/4$. Black-Scholes is given by (2), HS by (22), HS(0) by (52), HS(1) by (54) and HS(2) by (56).

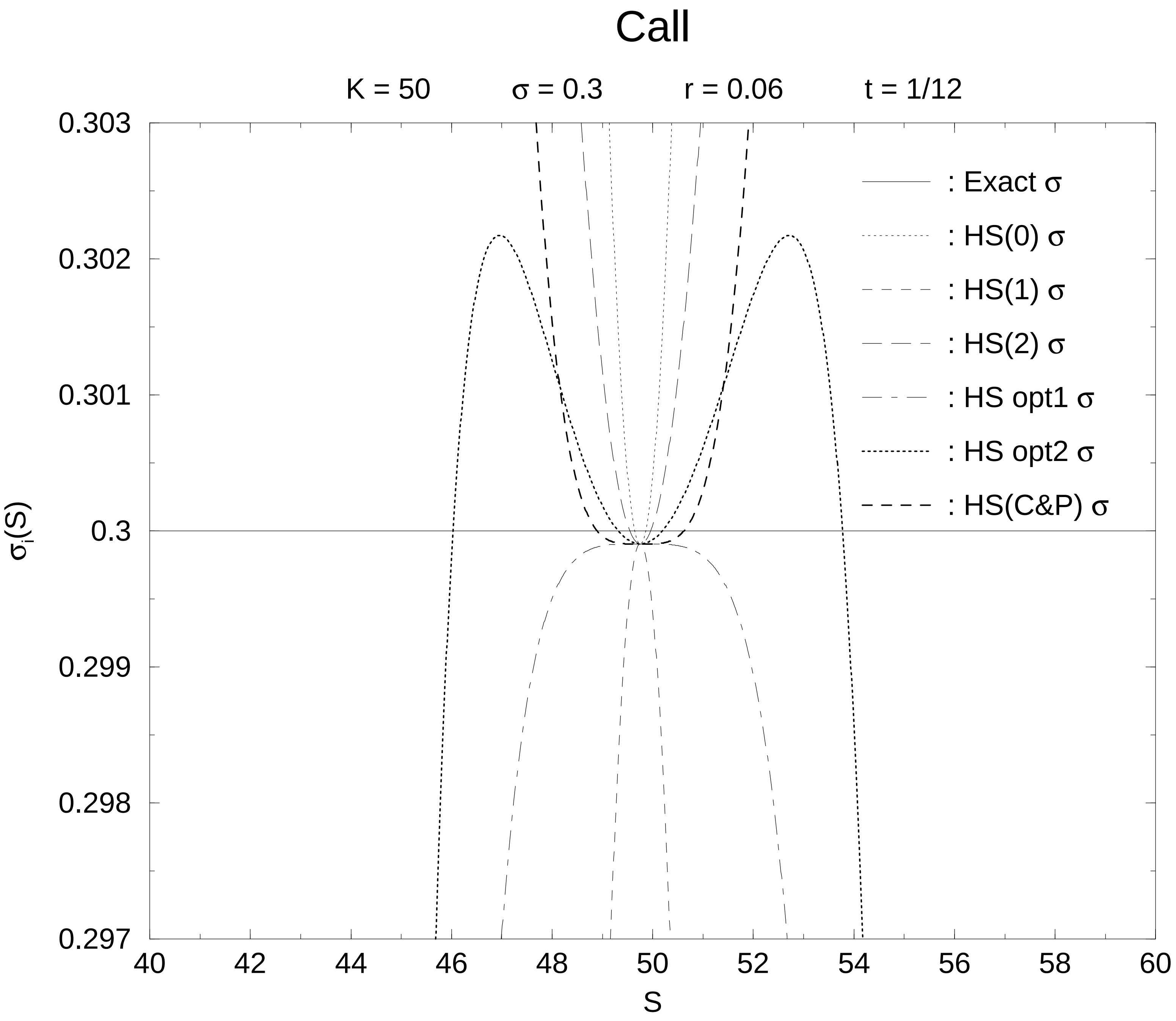

Call
K = 50
σ = 0.3
r = 0.06
t = 1/4
: Black−Scholes
: HS
: HS(0)
: HS(1)
: HS(2)
C(S)
S
60
40
20
0
−20
0
20
40
60
80
100

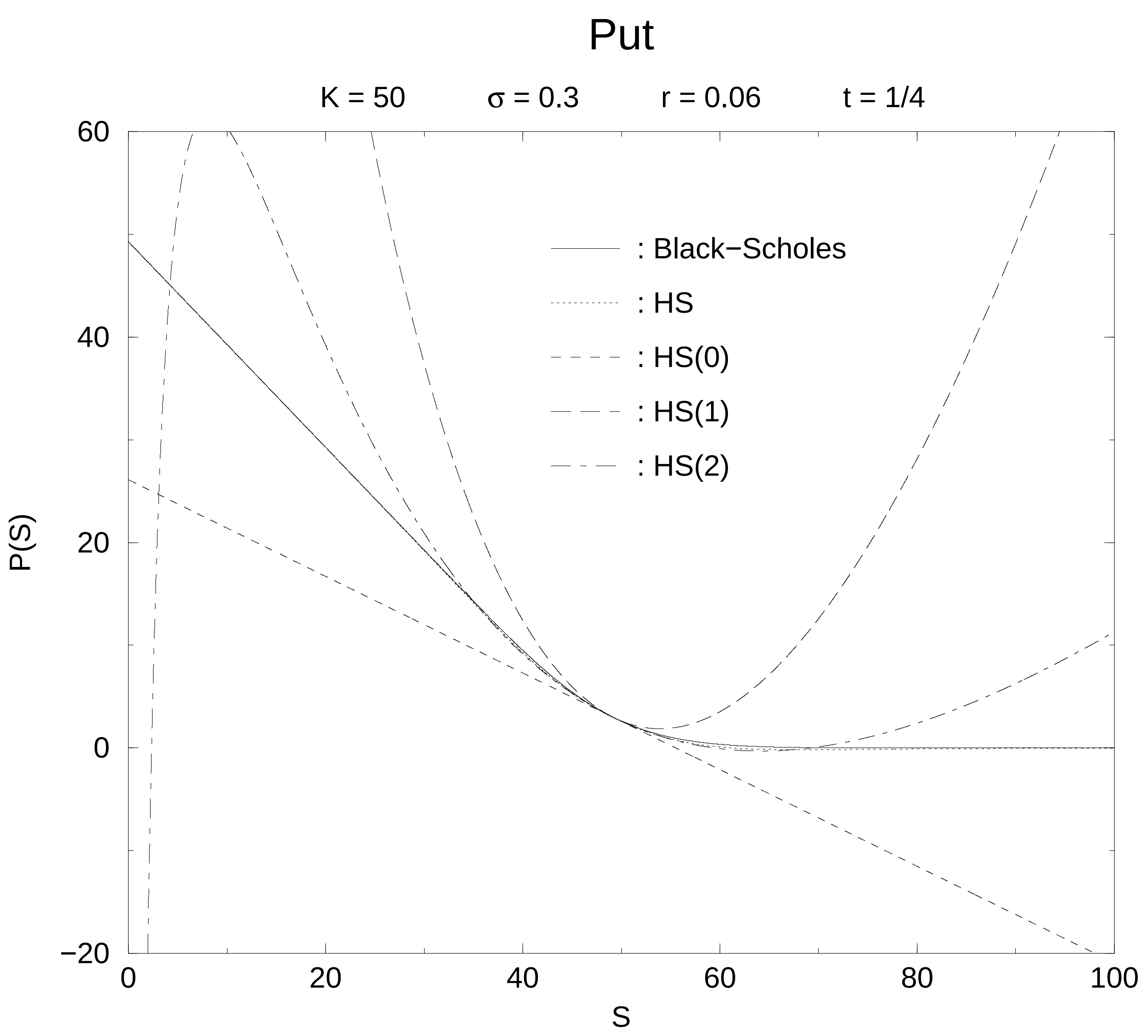

Call
K = 50
σ = 0.3
r = 0.06
t = 1/12
: Black−Scholes
: HS
: HS(0)
: HS(1)
: HS(2)
C(S)
S
60
40
20
0
−20
0
20
40
60
80
100


Call
K = 50
σ = 0.3
r = 0.06
t = 1/4
: Exact σ
: BS σ
: HS(0) σ
: HS(1) σ
: HS(2) σ
: HS opt2 σ
: HS(C&P) σ
σi(S)
S
2
1.6
1.2
0.8
0.4
0
0
20
40
60
80
100


Call
K = 50
σ = 0.3
r = 0.06
t = 1/12
: Exact σ
: BS σ
: HS(0) σ
: HS(1) σ
: HS(2) σ
: HS opt2 σ
: HS(C&P) σ
σi(S)
S
0
0.4
0.8
1.2
1.6
2
0
20
40
60
80
100


Call
K = 50
σ = 0.3
r = 0.06
t = 1/4
0.303
0.302
0.301
0.3
0.299
0.298
0.297
$\sigma_i(S)$
40
42
44
46
48
50
52
54
56
58
60
S
: Exact σ
: HS(0) σ
: HS(1) σ
: HS(2) σ
: HS opt1 σ
: HS opt2 σ
: HS(C&P) σ


Call
K = 50
σ = 0.3
r = 0.06
t = 1/12
: Exact σ
: HS(0) σ
: HS(1) σ
: HS(2) σ
: HS opt1 σ
: HS opt2 σ
: HS(C&P) σ
σi(S)
S
0.303
0.302
0.301
0.3
0.299
0.298
0.297
40
42
44
46
48
50
52
54
56
58
60



Put
K = 50
σ = 0.3
r = 0.06
t = 1/4
: Black−Scholes
: HS
: HS(0)
: HS(1)
: HS(2)
P(S)
S
60
40
20
0
−20
0
20
40
60
80
100