What could potentially aggravate implicit volatility "smile" and "asymmetry"? — A note

George J. Jiang*

Schulich School of Business

York University

Canada

March 12, 2001

*George Jiang, Department of Finance, Schulich School of Business, York University, 4700 Keele Street, Toronto, ON., Canada M3J 1P3. Tel: (416) 736-2100 Ext. 33302, Fax: (416) 736 5687. E-mail: gjiang@schulich.yorku.ca. George Jiang is also a SOM research fellow of the Faculty of Business and Economics at the University of Groningen in The Netherlands. The author wishes to thank Thierry Roncalli for helpful comments and suggestions. The usual disclaimer applies.

What could potentially aggravate implicit volatility "smile" and "asymmetry"? — A note

Abstract

This research note shows that the implied Black-Scholes volatility calculated using the *bisection algorithm* can have significant biases, which are more severe for in-the-money (ITM) options than for out-of-the-money (OTM) options. The biases are shown to have important implications as they could potentially aggravate the well-documented *smile* or *smirk* and *asymmetry* of implied Black-Scholes volatility for equity options. The findings caution the use of bisection algorithm for the calculation of Black-Scholes implied volatility. This research note also shows that the biases can be eliminated using the *optimization algorithm* which uses at least the first derivative of the objective function.

The implicit or implied volatility is defined as the value of the standard derivation of the underlying asset returns at given time that equates the option price calculated from appropriate option pricing formulas to the observed option price. In the case of the Black-Scholes (1973) model, the Black-Scholes implicit or implied volatility is defined as [1]

$$C_{BS}(S_t, K, \sigma^{imp}, T-t, r) = C^* \tag{1}$$

where C^* is the observed option price and

$$C_{BS}(S_t, K, \sigma, T-t, r) = S_t[\Phi(d) - e^{-x_t}\Phi(d - \sigma\sqrt{T-t})] \tag{2}$$

is the Black-Scholes formula for the European call option prices with $x_t = \ln(S_t/Ke^{-r(T-t)})$ and $d = x_t/\sigma\sqrt{T-t} + \sigma\sqrt{T-t}/2$, where S_t is the asset price at time t, K is the strike price of the option, $T(> t)$ is the expiration date of the option, r is the non-stochastic risk-free rate of return, and $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard normal distribution. In the above Black-Scholes formula, the risk-free rate r is assumed as constant and known, S_t is observed, the strike price K and expiration date T are specified in the

[1] In time series literature, both standard derivation σ and variance σ^2 are often referred as volatility.

option's contract, and the only unobserved parameter is the underlying volatility σ. Since $\partial C_{BS}/\partial\sigma = S_t\sqrt{T-t}\Phi'(d) > 0$, i.e. the sensitivity of option price with respect to the change of underlying volatility (often referred as option's *vega*) is strictly positive, $C_{BS}(\cdot)$ is a strictly increasing function of σ and there is a unique solution of σ^{imp} in equation (1).

In the continuous-time diffusion model framework, estimation of the diffusion parameter σ from underlying asset return time series is not as straightforward as it appears to be. Consider the following general Black-Scholes (1973) and Merton (1973) asset return model

$$dS_t/S_t = \mu(\cdot)dt + \sigma dW_t \tag{3}$$

in which the drift function $\mu(\cdot)$ can be non-constant but any arbitrary function of the asset price S_t and time t or even other economic variables, as pointed out in Lo and Wang (1995). It is noted that the drift term does not enter into the option pricing formula, as it is replaced by the risk-free rate of return in the risk-neutral process. However, the drift function $\mu(\cdot)$ and the diffusion coefficient σ are related to each other through the transition density function of the underlying asset return process and/or the marginal density function of a stationary underlying return process, thus the estimation of σ depends critically on the specification of the drift function $\mu(\cdot)$ given observations of underlying asset returns. For example, when the drift term is specified as a linear mean-reverting function, i.e. the asset return follows a trended Ornstein-Uhlenbeck process instead of a geometric Brownian motion process as in Black and Scholes (1973), the estimator of volatility σ^2 would be different than that of the Black-Scholes volatility based on the same set of asset return observations. For details, see Lo and Wang (1995). Only in the case that $\mu(\cdot) = \mu$, can the maximum likelihood estimators of $\alpha = \mu - \frac{1}{2}\sigma^2$ and σ^2 be obtained from $\hat{\alpha} = (\ln S_T - \ln S_0)/T, \hat{\sigma}^2 = \sum_{t=1}^{T}(\ln(S_t/S_{t-1}) - \hat{\alpha})^2/T$ based on a set of asset price observations $\{S_0, S_1, ..., S_T\}$. In fact, in pricing options the time series based definition of volatility is secondary and seldom used.

Since there is a one-to-one relationship between the volatility and the Black-Scholes option price as in equation (1), volatility can also be backed out from observed option price. By doing so, the specification of the drift term $\mu(\cdot)$ becomes irrelevant as it does not enter the option pricing formula. However, more often than not, $C^*_{\tau i}$ is a cross-section of option prices at time t for $\tau = T - t \in R_+$ and $i = 1, 2, ..., N$, there also will be a cross-section of implied volatility at time t, $\hat{\sigma}^{imp}_{\tau i}$, which is not necessarily constant. The Black-Scholes model imposes a flat term structure of volatility, i.e. the volatility is constant across both maturity and moneyness of options. If option prices in the market were confirmable with the Black-Scholes formula, all the Black-Scholes implied volatility corresponding to various options written on the same asset would coincide with the volatility parameter σ of the underlying asset. In reality this is not the case, and the Black-Scholes implied volatility heavily depends on the calendar time, the time to maturity, and the moneyness of the options. Many studies, e.g. Rubinstein (1985), use this approach to examine the cross-sectional pricing errors of the Black-Scholes model based on the implied volatility from observed option prices using equation (1). Empirical studies have found various patterns of the implied volatility across different strike prices and maturities. The price distortions, well-known to practitioners, are usually documented in the empirical literature under the terminology of the *smile* or *smirk* effect, referring to the U-shaped or tilted U-shaped pattern of implied volatilities across different strike prices. More specifically, the following stylized facts are extensively documented in the literature (see for instance Rubinstein (1985), Clewlow and Xu (1993), Taylor and Xu (1993)) for the implied Black-Scholes volatility: (i) The U-shaped pattern of implied volatility as a function of moneyness has its minimum centered at near-the-money options; (ii) The volatility smile is often but not always symmetric as a function of moneyness; and (iii) The amplitude of the smile or smirk increases quickly when time to maturity decreases. Indeed, for short maturity options the smile or smirk effect is very pronounced while it almost completely disappears for longer maturity options.

However, the inversion of implied volatility as a function of the observed option price $\sigma^{imp} = \sigma^{imp}(C^*)$ has no explicit analytical form, thus calculation of the implied volatility in general relies on numerical procedures. The commonly used methods in solving equation (1) are the *bisection algorithm* which searches the value of σ^{imp} through iteration and the *optimization algorithm* which minimizes the squared distance between the model price and observed price. We refer the former method as Method I and the latter as Method II.

Method I: bisection algorithm

Exploiting the fact that $C_{BS}(\cdot)$ is a strictly monotonic function of σ, the iterative bisection algorithm can be employed to solve equation (1). In this case, the first step amounts to choosing the values for a and b such that

$$C_{BS}(S_t, K, a, T-t, r) \leq C^* \tag{4}$$

and

$$C_{BS}(S_t, K, b, T-t, r) \geq C^* \tag{5}$$

The next step is to evaluate $C_{BS}(\cdot)$ at $c = (a+b)/2$, if $C_{BS}(S_t, K, c, T-t, r) \leq C^*$, then replace a by $(a+b)/2$ and continue the iteration. Otherwise, replace b by $(a+b)/2$ and continue the iteration. The stopping rule is if $|a-b| \leq \epsilon$, a pre-set permissible error, then

$\hat{\sigma}^{imp} = (a+b)/2$ is the solution. This method is used in Roncalli (1997) whose GAUSS codes are included in the American University GAUSS archives.

Method II: optimization algorithm

Also exploiting the fact that C_{BS} is a strictly monotonic function of σ, the following optimization can be employed to solve equation (1),

$$\hat{\sigma}^{imp} = \arg\min_{\sigma^{imp}} (C_{BS}(S_t, K, \sigma^{imp}, T-t, r) - C^*)^2 \tag{6}$$

which has a unique solution for σ^{imp}. The above minimization problem is rather simple and can be solved using any available optimization method.

The difference between these two methods is that Method I only involves the evaluation of the functional values and requires much less computational time, while method II involves the evaluation of the first derivative and sometimes second derivative (depending on the algorithm) of the objective function and requires more computational time. In this note we illustrate that the first method can lead to considerable biases of implied volatility, while the second leads to very accurate values of implied volatility. The program used in this note for the first method is from Roncalli (1997), which is posted in the internet with the address given in the reference, and the program for the second method is written by the author, both in GAUSS language. The convergence limit ϵ in the first method is set as 10^{-5} and the initial interval is set as $[a, b] = [10^{-8}, 0.1]$. The results are shown to be robust to the change of convergence limit and initial interval around these values. The optimization problem in the second method is solved using the procedure "OPTMUM" of GAUSS 3.2 with the default optimization method BFGS (i.e. the Broyden, Fletcher, Goldfab and Shanno method, see e.g. Broyden, 1965) which uses both first and second derivative information. Alternative optimization methods, such as the Newton-Raphson method which uses only the first derivative of the objective function, are shown to work also very well. The convergence is indicated by the return codes (0 indicates normal convergence). The starting value in method II is set as $\sigma = 1\%$ which is different from the true value.

For numerical analysis, we set the asset price as $S_t = \$40.00$, strike price as $K = [\$30, \$50]$, maturity date as $T - t = 21, 63, 126, 252$ business days (or 1-month, 3-month, 6-month, and 1-year respectively), standard derivation of daily continuously-compounded asset returns as $\sigma = 0.02$ or 2%, and the annual continuously-compounded risk-free rate as $r = 5\%$. Similar numbers are used in Lo and Wang (1995) for numerical analysis of option pricing models. Figure 1 shows the implied volatility calculated using both methods from the theoretical option prices generated by the Black-Scholes model. The horizontal axis represents moneyness of the options, which is measured by $\ln(S_t/K)$. It is obvious that Method

II yields very accurate valuation of implied volatility and the term structure of volatility is flat across both moneyness and maturity as assumed in the Black-Scholes model. However, Method I results in considerable biases of implied volatility, with the implied volatility of short-term in the money (ITM) options in particular overvalued. Thus the implied volatility appears to be asymmetric and skewed to the left. The implied volatility curve of short-term options has a more pronounced curvature, while that of long-term options is relatively flat. The asymmetric implied volatility curve calculated using method I happens to be agreeable with the well-documented implied Black-Scholes volatility “*smile*" or “*smirk*".

Figure 2 shows the implied volatility calculated using both methods from the theoretical option prices generated by the Merton (1976) jump-diffusion model. The jump-diffusion model in Merton (1976) assumes that asset return follows a mixture of continuous diffusion path and discontinuous jump path, i.e.

$$dS_t/S_t = (\alpha(\cdot) - \lambda\alpha_0)dt + \sigma dW_t + (Y_t - 1)dQ_t(\lambda) \tag{7}$$

where

$\alpha(\cdot)$ —the instantaneous expected return on the asset;

σ^2 — the instantaneous volatility of the asset's return conditional on no arrivals of important new information (i.e. the Poisson jump event does not occur);

$Q_t(\lambda)$ — a Poisson counting process which is assumed to be i.i.d. over time, λ is the mean number of jumps per unit of time, i.e. the intensity parameter of the Poisson distribution with Prob$(dQ_t(\lambda) = 1) = \lambda dt$, Prob$(dQ_t(\lambda) = 0) = 1 - \lambda dt$;

$Y_t - 1$ — the random jump size $(Y_t \geq 0)$ representing the percentage change of asset price if the Poisson event occurs, $\int_0^t (Y_\tau - 1)dq_\lambda(\tau)$ is a compound Poisson process, and α_0 is the expectation of the relative jump size, i.e. $\alpha_0 = E[Y_t - 1]$;

$dQ_t(\lambda)$, dW_t — assumed to be statistically independent.

Merton (1976) shows that, when Y_t follows a log-normal distribution, i.e. $\ln Y_t \sim$ iid $N(\ln(1+\alpha_0) - \frac{1}{2}\nu^2, \nu^2)$, thus $Y(n)$ has a log-normal distribution with the variance of logarithm of $Y(n)$, $Var[\ln Y(n)] = \nu^2 n$, and $E_{Y(n)}[Y(n)] = (1 + \alpha_0)^n$, a closed-form solution of the European call option price is given by

$$C_M(S_t, t) = \sum_{n=0}^{\infty} \frac{e^{-\lambda'\tau}(\lambda'\tau)^n}{n!} C_{BS}(S_t, K, \nu_n, T - t, \gamma_n) \tag{8}$$

where $\lambda' = \lambda(1+\alpha_0)$, $\nu_n^2 = \sigma^2 + n\nu^2/\tau$, $\gamma_n = r - \lambda\alpha_0 + n\ln(1+\alpha_0)/\tau$, and $\tau = T - t$. The option price is simply the weighted sum of the Black-Scholes price conditional on knowing that exactly n Poisson jumps will occur during the life of the option with each weight being the probability that a Poisson random variable with intensity $\lambda'\tau$ will take on the value n. In Figure 2, the asset price, strike price, maturity dates of the option and the risk-free rate of return are set as the same as in Figure 1. The jump and volatility parameter values are set as $\sigma^2 = 0.75 \times (0.02)^2, \lambda = 1/4, \alpha_0 = 0.0, \nu^2 = \times(0.02)^2$, i.e. the jump intensity is 25% and the expectation of jump size is zero. Since the asset return distribution has fatter tails due to the mixture of normal distributions and is symmetric due to the zero expectation of jump size, the implied volatility curve is expected to have a smile shape and be symmetric. When calculating the jump diffusion option prices, we truncate the summation in (8) from $n = 0$ to $n = 50$. As shown in Ball and Torous (1985) and verified in our calculation, the error due to this truncation is negligible. As indicated by the convergence codes and the symmetry of asset return distribution ($\alpha_0 = 0.0$), Method II again gives very accurate values of implied volatility. The implied volatility is essentially symmetric around the at-the-money (ATM) options and exhibits obvious U-shaped pattern or *smile*. The implied volatility curve of short-term options exhibits very pronounced smile, while that of long-term options is almost flat. However, Method I again generates significant upward biases of implied volatility, in particular for the short-term deep ITM options. As a result of the biases, the implied volatility appears to be asymmetric, skewed to the left, and has its minimum at in-the-money (ITM) options. Compared to the true implied volatility curves, the implied volatility curves exhibit much more pronounced smiles for short-maturity options and smirks for longer-maturity options. The asymmetry of the implied volatility curves appears to suggest that there might be a negative skewness, or negative expected jump, in the underlying asset returns. And again, the asymmetric implied volatility curve calculated using method I suggests that the deep ITM call options (equivalently the deep OTM put options by put-call parity) are priced at significant premiums relative to the deep OTM call options. Such numerical biases thus potentially have very important implications since the smiles or smirks and asymmetry of implied volatility are actually also the well-documented stylized facts in the literature for equity option prices. These biases caution the use of bisection algorithm as they could potentially aggravate the “smile" or “smirk" and the asymmetry of implied Black-Scholes volatility.

The problem for method I is a numerical problem of accuracy since it only involves the evaluation of the option pricing formula in the right-hand side of equation (1), i.e. the evaluation of the CDF of the standard normal distribution. A known fact about the normal distribution is that it has very thin tails, its numerical functional values over the thin tails would be virtually constant and thus have poor numerical accuracy, see Figure 3. The poor numerical accuracy in the procedure of method I can be further confirmed by the following calculations. We first calculate Black-Scholes option prices for deep ITM options with the input of different volatility values σ. The Black-Scholes option pricing formula generates almost the same option prices C_{BS} for a wide range of different volatility input σ, as the plots of CDF $\Phi(\cdot)$ against σ suggest that they are extremely flat for different inputs of σ. We then use Method I to calculate implied volatility from these option prices, it turns out that the implied volatilities for these deep ITM options are almost the same even though we know that the underlying Black-Scholes volatilities are actually different. When the iterative search procedure of the bisection algorithm starts with evaluating whether $C_{BS}(S_t, K, c, T-t, r) \leq C^*$ holds, where $c = (a+b)/2$ (i.e. choosing the high-range value of σ which satisfies the numerical equality (1)), there tends to be an over-valuation of the implied volatility. Such biases are difficult to eliminate by the bisection algorithm itself due to the limited precision of computers. Intuitively, due to the very thin tails of the normal distribution, the prices of short-term deep ITM options are not sensitive to the change of the level of volatility. That is, *vega* $= \partial C_{BS}/\partial\sigma = S_t\sqrt{T-t}\Phi'(d)$ is small when both $T-t$ and d are small. As in such cases, the asset price has 0^+ probability dropping below the strike price K for quite a range of different values of σ. (It should be noted that based on empirical asset return distributions with negative skewness and excess kurtosis, such probability may not be zero but positive, and yet such options can be actively traded in the market.) Method II, on the other hand, uses the first derivative and in some cases the second derivative (depends on the algorithm) in the optimization procedure. The first and second order conditions of the optimization problem in (1) are respectively $2S_t\sqrt{T-t}\Phi'(d)(C_{BS}-C^*)$ and $2S_t\sqrt{T-t}\Phi''(d)(C_{BS}-C^*) + 2(S_t\sqrt{T-t}\Phi'(d))^2$. From the plots of functional values of the normal CDF $\Phi(d)$, the normal PDF $\Phi'(d)$, and the first derivative of the normal PDF $\Phi''(d)$ as shown in Figure 3, it is obvious that $\Phi''(d)$ is more sensitive to the change of d than $\Phi'(d)$ is, and further $\Phi'(d)$ is more sensitive to the change of d than $\Phi(d)$ is. Noticeably $\Phi(d)$ is virtually flat for high values of d, while $\Phi'(d)$ and $\Phi''(d)$ are changing with d even for very high values of d. The use of first order condition and in some cases the second order condition as well thus leads to a more accurate solution of the implied volatility.

The numerical biases of implied volatility due to method I (the bisection algorithm) are not trivial for at least the following two reasons. Firstly, as we have noted, the above reported biases of implied volatility calculated using bisection algorithm happen to coincide with the well-documented empirical stylized facts about equity option prices. Due to the biases, the documented smiles or smirks and asymmetry might be, to a certain extent, aggravated by the numerical biases. Secondly, the biases are found to be more severe for ITM options than for OTM options. As well-known, it is the correction of the pricing errors for relatively expensive options, namely the long-term and ITM options which are of more economic significance, that presents one of the most severe challenges in option pricing theory. Moreover, the biases of implied volatility may provide misleading information for developing alternative option pricing models. Therefore, it is important to implement the accurate algorithm in computing implied volatility.

References

Ball, C.A. and W.N. Torous (1985), "On jumps in common stock prices and their impact on call option pricing", *Journal of Finance*, **40**, 155–173.

Black, F. and M. Scholes (1973), "The pricing of options and corporate liabilities", *Journal of Political Economy*, **81**, 637–654.

Broyden, G. (1965), "A class of methods for solving nonlinear simultaneous equations", *Mathematics of Computation*, **19**, 577–593.

Clewlow, L. and X. Xu (1993), "The dynamics of stochastic volatility" [Discussion paper, University of Warwick].

Lo, A. and J. Wang (1995), "Implementing option pricing formula when asset returns are predictable", *Journal of Finance*, **Vol. L(1)**, 87–129.

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

Merton, R.C. (1976a), "Option pricing when underlying stock returns are discontinuous", *Journal of Financial Economics*, **3**, 125–144.

Roncalli, T. (1997), "Gauss et al finance" [L.A.R.E., Université Montesquieu-Bordeaux IV, http:// gurukul.ucc.american.edu/econ/gaussres/utilitys/finance.src].

Rubinstein, M. (1985), "Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active cboe options classes from august 23, 1976 through august 31, 1978", *Journal of Finance*, **40**, 455–480.

Taylor, S. J. and X. Xu (1993), "The magnitude of implied volatility smiles: Theory and empirical evidence for exchange rates" [Discussion paper, University of Warwick].

Figure 1: Implied Volatility of the Black-Scholes Model using Different Algorithms

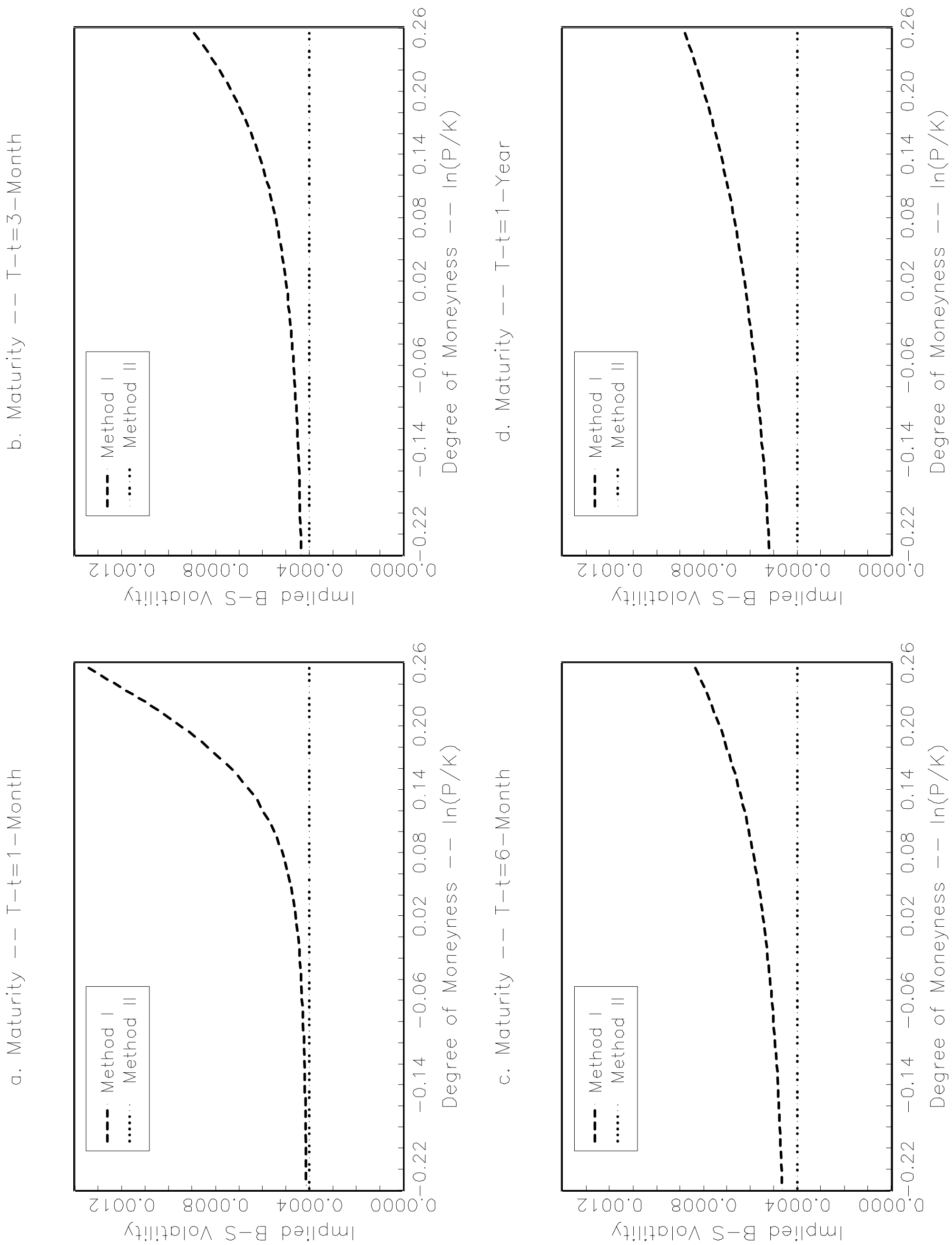

Figure 2: Implied Volatility of the Merton Jump-Diffusion Model using Different Algorithms

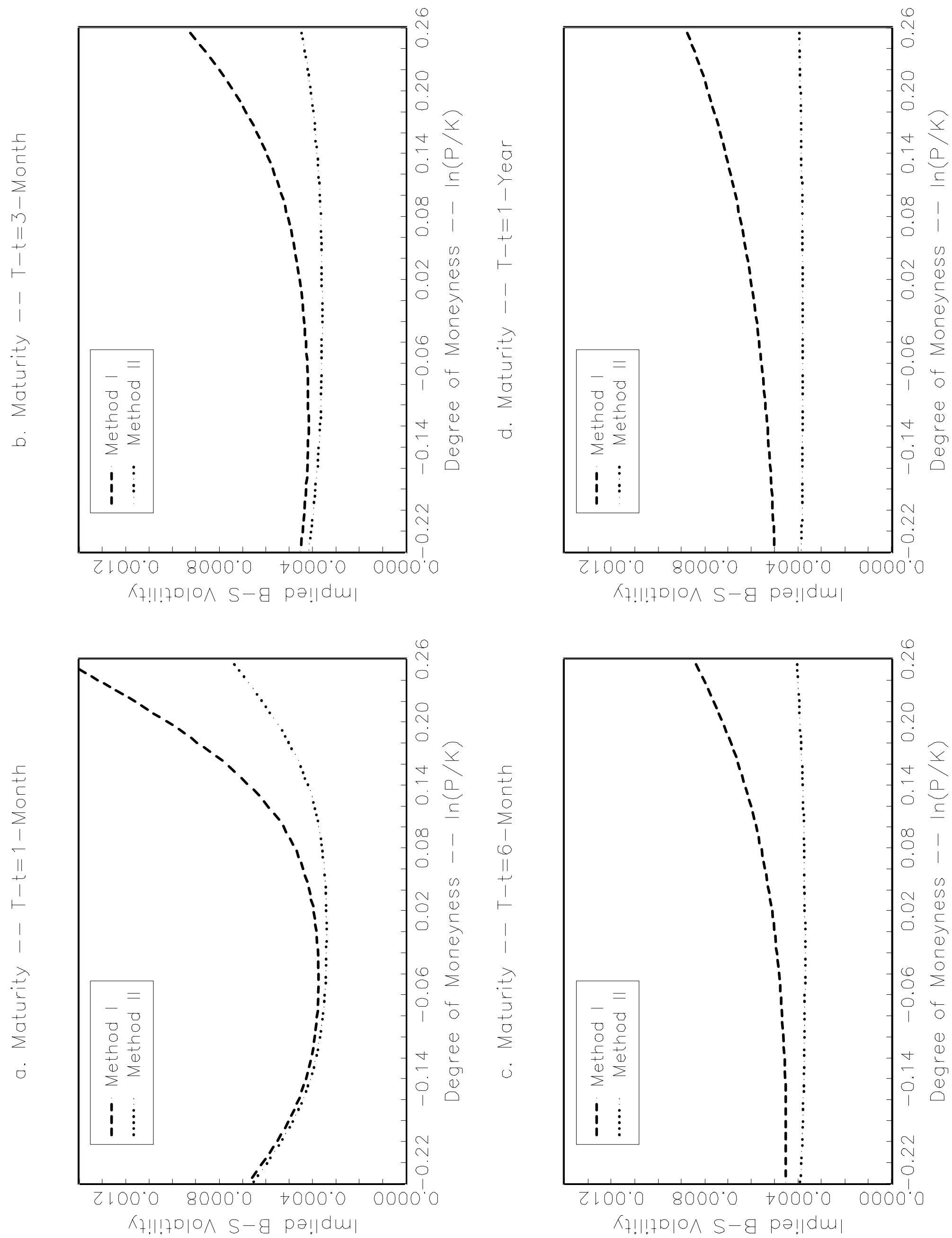

Figure 3: Functional Values of the Normal CDF, PDF and the First Derivative of PDF