SCCS-260
published in The Fourth Symposium on the Frontiers of Massively Parallel Computation
October 19-21, 1992, McLean, Virginia
IEEE Computer Society, NASA Goddard Space Flight Center

A Large Scale Comparison of Option Pricing Models with Historical Market Data*

Kim Mills, Michael Vinson, and Gang Cheng
Northeast Parallel Architectures Center at Syracuse University
Syracuse, NY 13244-4100
kim@npac.syr.edu

Abstract

A set of stock option pricing models are implemented on the Connection Machine-2 and the DECmpp-12000 to compare model prices and historical market data. Improved models, which incorporate stochastic volatility with American call generally have smaller pricing errors than simpler models which are based on constant volatility and European call. In a refinement of the comparison between model and market prices, a figure of merit based on the bid/ask spread in the market, and the use of optimization techniques for model parameter estimation, are evaluated. Optimization appears to hold great promise for improving the accuracy of existing pricing models, especially for stocks which are difficult to price with conventional models.

1 Introduction

Following the opening of the first organized options exchange in April, 1973 by the Chicago Board of Options Exchange, rapid growth in option trading has been accompanied by the development of option pricing theory and modeling. While there are many types of options, all option contracts are based on puts, calls, and an underlying asset (a stock or an index of stocks). The owner of a call option contract has a right but not the obligation to purchase shares of the asset for an agreed upon exercise or striking price, for a fixed period of time [2]. European option contracts can be exercised only at maturity, while American contracts can be exercised at any time during the life of the contract. Option traders include both speculators and financial managers. Speculators are attracted to the options market because of the potential for high profits. Considerably less capital is required to participate in the options market than the stock market. Financial managers participate in the options market to hedge risk in their portfolios.

The variance of asset price over time (defined as volatility) is a key parameter in any calculation of option prices. Since the introduction of a constant volatility, European pricing model (Black-Scholes) [1], finance researchers have sought improved methods to price options with stochastic volatility and American contracts.

A schematic view of the path of stock price over time is illustrated in Figure 1. Elements of the model

*We gratefully acknowledge support for this study from the Office of the Vice President for Research and Computing at Syracuse University, and Corporate Partnership funding from Digital Equipment Corporation.

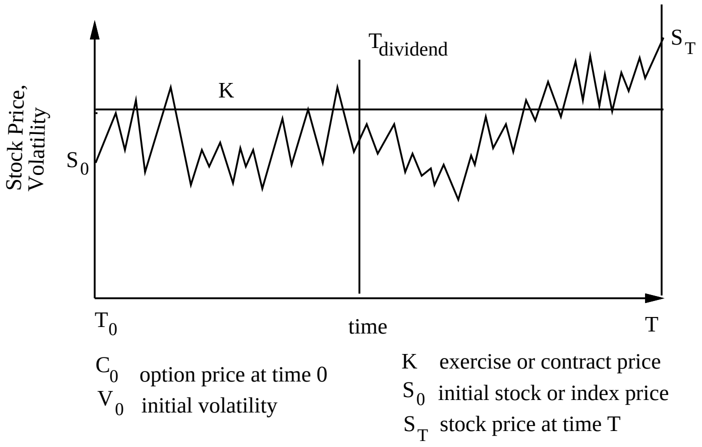

Figure 1: Schematic view of stock over life of option contract

include initial stock price, the call price, the exercise price, the time of dividend, as well as model parameters which cannot be directly observed but must be estimated from market information. These parameters include volatility of underlying asset, variance of the volatility, and correlation between asset price and volatility. In general, the time just prior to dividend payout is the only time that a call option is exercised before maturity.

This project is part of a program at the Northeast Parallel Architectures Center (NPAC) to develop applications of parallel computing in industry, and is the result of a collaboration with the School of Management at Syracuse University. Our purpose here is to report the results of a comparison of a set of option pricing models and historical market data. Performing this comparison requires high performance computing. In a related study we examine data distribution, load balancing, and communication issues and their effect on performance of option pricing models on the Connection Machine-2 and the DECmpp-12000 [7].

In this comparison, we observed smaller errors in pricing models incorporating stochastic volatility and American call than for models based on constant volatility and European call. In a refinement of our comparison between model and market prices, we used optimization techniques to estimate model parameters and devised a figure of merit based on the bid/ask spread in the market to summarize model performance. Optimization appears to hold great promise in improving the performance of existing pricing models. Current, ongoing work includes developing a simple trading strategy to assess model performance in terms of market profitability.

2 Option pricing models

The Black-Scholes option pricing model was first published in 1973 [1] with the opening of the Chicago Board of Options Exchange, and remains commonly used. This model assumes constant volatility and European pricing (exercise only at maturity), and is the least sophisticated model considered in this study. Black and Scholes [1] derived a nonstochastic equation for call price that can be solved analytically. Many of the models that follow the Black-Scholes model incorporate methods for treating volatility as a stochastic process.

Monte Carlo models are the conventional standard of comparison for option pricing models. The Monte Carlo method allows us to directly incorporate volatility and stock price change as stochastic processes, and parallelizes very easily. While generally accepted to provide the most accurate pricing estimates, Monte Carlo models remain too computationally intensive to be used other than for research purposes.

Binomial approximation models allow us to incorporate stochastic volatility and American call, and are computationally far more efficient than Monte Carlo simulation. In a previous, related study, Finucane [4] compared a set of Monte Carlo simulation and binomial pricing models. Using a set of fixed input parameters (stock price, volatility, variance of volatility, correlation, stock price/exercise price ratio), binomial models were demonstrated to provide accurate approximations (within two standard errors) of the stochastic volatility price for the European and American Monte Carlo cases.

In this study, we evaluate the accuracy of binomial approximation models for pricing call options. We selected four pricing models, implemented these models in Fortran90 on the Connection Machine-2 and the DECmpp-12000, and performed a comparison between model and historical market prices. The four models in our market comparisons are:

Model 1. Black-Scholes model (constant volatility, European call)

Model 2. Binomial approximation with constant volatility, and American call

Model 3. Binomial approximation with stochastic volatility, and European call

Model 4. Binomial approximation with stochastic volatility, and American call

Following [3, 5, 4], we briefly summarize the equations describing the continuous time movement of stock price and volatility (variance of stock price) over the life of an option contract. Discretizing these processes within the binomial lattice is based on an assumption that stock price and volatility follow a continuous drift. The binomial model is used to derive a distribution of stock prices at time of maturity.

Volatility, σ, and stock price, S, follow stochastic processes represented as

$$\frac{d\sigma^2}{\sigma^2} = \mu_\sigma dt + \xi d\widetilde{W} \tag{1}$$

$$\frac{dS}{S} = \mu_s dt + \sigma d\widetilde{Z} \tag{2}$$

where $\widetilde{W}$ and $\widetilde{Z}$ are standard Weiner processes with correlation ρ, μ_σ is the drift of the variance process and μ_s is the drift of stock price (both constants) and ξ is the volatility of the variance (not directly observed, but estimated from data). Weiner processes generate continuous paths that are in constant motion no matter how small the time step.

Binomial approximation models represent the continuous time processes described above as a lattice of discrete up/down movements in stock price and volatility. For example, the magnitude of the increase (u) or decrease (d) in variance for a given time period is as

$$u = e^{(\mu_\sigma - \xi^2/2)\Delta t + \xi\sqrt{\Delta t}} \tag{3}$$

$$d = e^{(\mu_\sigma - \xi^2/2)\Delta t + \xi\sqrt{\Delta t}} \tag{4}$$

with the probability of an increase or decrease being equally likely. With the introduction of correlation, ρ, the variance of stock price after i periods with j upward movements and $i-j$ downward movements is then defined as

$$\sigma^2 = \left(\sigma_{0,0}^2\right) u(\rho)^i d(\rho)^{i-j} \tag{5}$$

In the limit, as Δt approaches zero, the binomial process approaches the continuous time process.

The magnitude of increases (U) and decreases (D) within the stock price are then defined as

$$U_{i,j} = e^{(r_f - \sigma_{i,j}^2/2)\Delta t + \sigma_{i,j}\Delta t} \tag{6}$$

$$D_{i,j} = e^{(r_f - \sigma_{i,j}^2/2)\Delta t - \sigma_{i,j}\Delta t} \tag{7}$$

American options incorporate early exercise, which means that the option can be exercised at any time during the life of the contract. Pricing American option contracts with the binomial model requires tracking price movements within the lattice from the time of early exercise (dividend payout) to contract maturity. We use American pricing, but do not describe details of the model implementation in this paper.

3 Implementation of the binomial approximation model

Binomial models provide a numerical procedure for approximating the stochastic processes of stock price change over time. A binomial lattice is illustrated in Figure 2 showing asset price or volatility of price in the vertical axis and time in the horizontal axis. Important elements of the model include initial price (S_0) and volatility (σ_0) or (V_0), time of dividend payout (t_{div}), the $2^{t_{\mathrm{div}}}$ nodes at time of the dividend where t_{div} ranges over values 1 to $T-1$, and the 2^T nodes at terminal time T. A single option price C_0, is estimated from a weighted average of the 2^T prices at time T and discounting to the present time T_0.

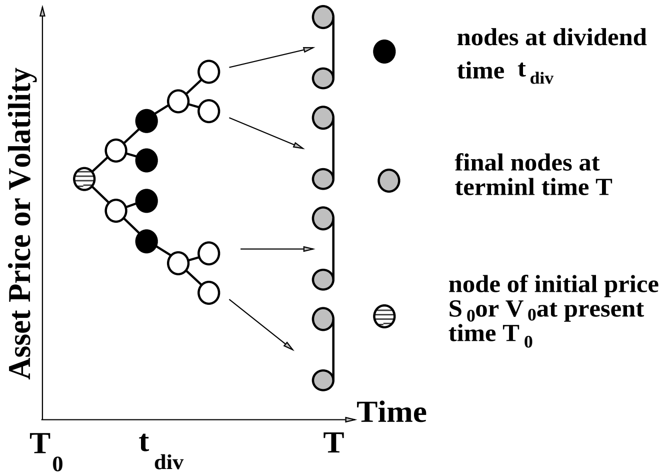

Figure 2: Structure of the binomial lattice

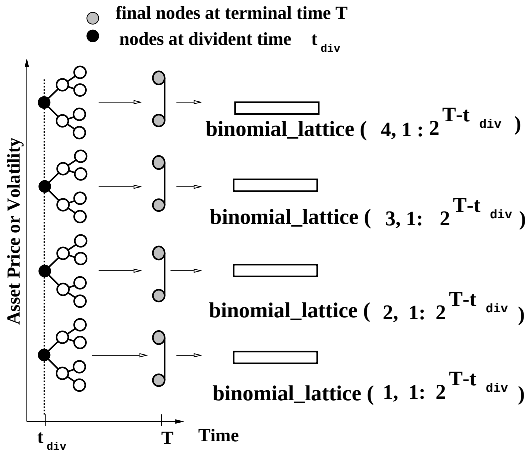

Figure 3: Binomial lattice expressed as a two-dimensional array

We designate the time steps in our model from 1 to t_{div} as stage 1 of the model, and timesteps from t_{div} to maturity T as stage 2 of the model. This breakdown of the American pricing model allows us to track price movements after dividend payout and determine percentages of early exercise.

Figure 3 illustrates the $2^{t_{\text{div}}}$ nodes in the binomial lattice at time of dividend. The value of t_{div} ranges from 1 to $T-1$ and defines the shape of the two-dimensional Fortran array $(1:2^{t_{\text{div}}}, 1:2^{T-t_{\text{div}}})$. The value t_{div} comes from market information (each option record has its own value t_{div}) and is not accessible to the model until run-time, requiring dynamically allocated arrays.

At the close of stage 1 in our model, there are $2^{t_{\text{div}}}$ nodes in the lattice. After dividend payout, and the onset of stage 2 of the model, up/down movements of price (and volatility) for each node are represented by a subtree of the lattice and expressed in the second dimension of the array of size $2^{T-t_{\text{div}}}$. As illustrated in Figure 3, when $t_{\text{div}} = 2$, there are $2^{t_{\text{div}}}$ or 4 rows in the two-dimensional array. After dividend payout, stage 2 of the model, further up/down moves of price and volatility are expressed in the 2^{T-2} columns of the two-dimensional array.

We run the binomial model for T= 17 time steps or periods. At each time step T, there are 2^T points in the binomial lattice. This model size is consistent with a related study [4], we also found little improvement in model accuracy with model sizes of T= 18, 19 and 20. We express the binomial lattice in both one and two-dimensional arrays, and use the Fortran90 intrinsic function eoshift to perform repeated nearest-neighbor communication within the array.

Using this approach, we evolve binomial lattices of volatility and price over 17 periods. In each period, the number of nodes in the lattice doubles, representing the up/down movements of volatility and price over time.

Although the communication requirements of this model are significant, the model requires only nearest neighbor communication along one axis. This feature allows us to express the the two-dimensional structure of the binomial lattice (volatility x time, and price x time) in one-dimensional Fortran arrays. We describe implementation of one and two-dimensional models on the Connection Machine-2 and DECmpp-12000 in a related study [7]. Data distribution, load balancing, and communication issues have an important influence on model run time.

4 Comparison of market and model prices

We obtained market data from the Chicago Board of Options Exchange for the period 1988-1990, and in our initial tests, used one-month records of option trades from January, 1988 for a set of 13 stocks. Options are a high-volume instrument, each one month data set consists of individual trades ranging in size from three to ten thousand trades.

We define the market price as the average of the prevailing bid and asking prices for each trade record. From our set of pricing models, we calculate four models prices for each trade and compare model with market prices. Before running the pricing models, we must first estimate various model parameters. Volatility σ, the most important parameter in all of the models, is not directly observable. The same is true of ξ, the variance of σ, and its correlation with the stock price ρ. The techniques used to estimate these parameters has a direct impact on the data comparison.

We begin with a simple method for model parameter estimation. At the beginning of each half hour interval, we select an option with an exercise price closest to the stock price, and the shortest expiration time. We compute four estimates of market volatility, termed the implied volatility, by numerically inverting the four models for the selected option record. These implied values of volatility are then used as input to the models to price the remaining options in that half hour interval. To estimate ξ and ρ, which are assumed in this simple method not to vary over time, we average the half hour implied volatilities for each day, and compute the variance and correlation of these daily averages over the month long market record. We compare market and model prices by reporting RMS errors.

In our preliminary comparison, we examined the performance of four pricing models, using this simple method of parameter estimation, over the one month period of January, 1988. The following list identifies the individual stocks used in this comparison: Bristol Myers Squibb (bmy), Chrysler Corp. (c), Eastman Kodak (ek), Ford Motor Corp. (f), General Electric (ge), Hewlett Packard (hwp), International Business Machines (ibm), American Telephone & Telegraph (t), Texas Instruments (txn), Walmart (wmt), and Xerox (xrx). Figure 4 represents model performance as a RMS error between model and market price. For all stocks, the least sophisticated model, model 1 Black-Scholes with constant volatility and European call, has the largest errors. Our most sophisticated model, model 4 binomial model with stochastic volatility and American call, tends to have the smallest pricing errors. Pricing errors also tend to vary by stock. For example, AT&T (t) seems to be harder to price than Ford Motors (f) for this period. Although not shown here, we observed that the more sophisticated binomial model (stochastic volatility with American call) tends to perform better than the other models for options with the longest times to maturity (for example, greater than 60 days).

Figure 4: Results of preliminary market comparison

5 Refining the comparison of models and market data

In our initial evaluation of model accuracy, we used a simple method of model parameter estimation and expressed model errors as RMS errors. To refine this comparison of model and market prices, we developed a figure of merit to summarize model accuracy, and used optimization techniques to estimate model parameters.

The figure of merit is based on the bid/ask spread in the market. As the term implies, the bid/ask spread is the difference between prevailing bids by buyers, and asking prices by sellers for a given option. Our figure of merit defines the percentage of time that a model price falls within one bid/ask spread of the market price (defined as the average of the bid and ask price). The figure of merit provides a simple method for summarizing model accuracy in market terms.

In addition to the simple method of parameter estimation based on historical values, we investigated a more sophisticated approach using nonlinear optimization. In this scheme, half-hour volatilities are estimated as described above, and the parameters ξ and ρ are chosen so that they minimize the χ^2 error between the market prices and the model prices for each day. To do this, we define the χ^2 as

$$\chi^2(\xi, \rho, t) = \sum_{i=1}^{N_t} (P_i - M_i(\xi, \rho))^2 \tag{8}$$

where N_t is the number of records for day t, P_i is the market price of the i^{th} record, and $M_i(\xi, \rho)$ is the model price for the i^{th} record using parameters ξ and ρ. χ^2 is a nonlinear function of ξ and ρ and turns out to be a rather smooth function of the parameters. This allows us to use the downhill simplex method [6], a simple method of nonlinear optimization which works well for this application. Using optimization techniques, we find the parameters ξ and ρ that give the best possible fit to the data.

Estimating these parameters requires a great deal of computational effort. For a typical run, the downhill simplex method requires approximately 20 steps to converge to parameters with an accuracy of 10^{-3}. Each step requires the calculation of model prices and implied volatilities for all of the options in the given day. Typical data sets include 100 trades per day, 14

Figure 5: Figure of merit summarizing model performance

of which are numerically inverted to compute the implied volatilities. The numerical inversion requires, on average, about 10 price calculations. Thus to estimate the parameters for one day requires 20*(86 + 14*10) = 4520 option price calculations.

We used model parameter estimates based on optimization as input to model 4, the binomial model with stochastic volatility and American call. Figure 5 summarizes results for a subset of the 13 stocks in our previous comparison. Model numbers 1 through 4 correspond to the same four models used above. Models 5 and 6 are based on model 4, but use optimized parameter estimates for ξ and ρ. In general, optimization substantially improves the figure of merit summarizing model performance. Improvement in model accuracy with optimized parameters is greatest for IBM and Eastman Kodak stock in this period (January, 1988). IBM appears to be more difficult to price than other stocks in our sample, so we might expect optimization to make a difference. Eastman Kodak stock, however, is reasonably modeled without optimization without optimization, and further improved with optimization.

6 Discussion and conclusion

We used parallel models to perform a large scale comparison of option pricing models and historical market data. It is important to note that a small percentage improvement in model accuracy has important implications for this application. While our comparison was limited to one month of market data, our results suggest that improved pricing models, incorporating stochastic volatility and American call, are more accurate than simple models based on constant volatility and European call. Incorporating optimization techniques into option pricing appears to hold great promise.

This comparison required approximately 150 hours of 8K Connection Machine-2 or DECmpp-12000 time to perform. Based on speedup ratios observed in a related study [7], we estimate a similar comparison using sequential models running on a high speed workstation, such as a SUN4 or DECstation 5000 would require approximately 7,000 hours.

Current work related to this project includes further improvement of optimization techniques [8], and application of the models to longer time periods. In addition, we are developing a simple trading strategy to assess model accuracy in terms of market profitability. In this strategy, we use the models to identify under priced options in the market, buy and hold options for various holding periods, and track long term profitability of the various models.

In conclusion, this study demonstrates an application parallel computing in the finance industry. Parallel models are required for performing large scale comparisons between model and market prices. Parallel models are useful tools for developing new pricing models and applications of pricing models, such as pricing entire portfolios and devising hedging strategies under changing market conditions.

References

[1] F. Black, and M. Scholes. "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 1973, 637-59. 1973.

[2] T. Copeland, and J. Weston. "Financial Theory and Corporate Policy," *Chapter 8. Pricing Contingent Claims: Option Pricing Theory and Evidence.* pp. 240-298. Addison-Wesley. New York. 1986.

[3] J. Cox, S. Ross, and M. Rubinstein. "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, Vol. 7, pp. 229-63, 1979.

[4] T. Finucane, "Binomial Approximations of American Call Option Prices with Stochastic Volatilities," to be published in *Journal of Finance.* 1992.

[5] J. Hull, and A. White. "The Pricing of Options on Assets with Stochastic Volatilities," *The Journal of Finance.* Vol 42:2, pp. 281-300. 1987.

[6] S.L.S Jacoby, J.S. Kowalik, and J.T. Pizzo, *Iterative Methods for Nonlinear Optimization Problems.* Englewood Cliffs, N.J.: Prentice Hall, 1972.

[7] K. Mills, G. Cheng, M. Vinson, S. Ranka, G. Fox. "Software Issues and Performance of a Parallel Model for Stock Option Pricing." to be published in *The Fifth Australian Supercomputing Conference. Syracuse Center for Computational Science-273b*, 9 pps. 1992.

[8] M. Vinson, K. Mills, and G. Cheng. "Optimization Techniques for Option Pricing", *Syracuse Center for Computational Science-draft*, 1992.