Option Pricing Under a Normal Mixture Distribution Derived from the Markov Tree Model

Harish S. Bhat*† Nitesh Kumar*

April 21, 2012

Abstract

We examine a Markov tree (MT) model for option pricing in which the dynamics of the underlying asset are modeled by a non-IID process. We show that the discrete probability mass function of log returns generated by the tree is closely approximated by a continuous mixture of two normal distributions. Using this normal mixture distribution and risk-neutral pricing, we derive a closed-form expression for European call option prices. We also suggest a regression tree-based method for estimating three volatility parameters σ, σ^+, and σ^- required to apply the MT model. We apply the MT model to price call options on 89 non-dividend paying stocks from the S&P 500 index. For each stock symbol on a given day, we use the same parameters to price options across all strikes and expiries. Comparing against the Black-Scholes model, we find that the MT model's prices are closer to market prices.

1 Introduction

The Black-Scholes model for European call options assumes that the underlying asset follows a geometric Brownian motion: if S_t is the price of the underlying at time t, then $dS_t = \mu S_t dt + \sigma S_t dW_t$, where μ and σ are constants and W_t is a Brownian motion. It follows that the Black-Scholes model assumes normality of daily log returns and independence of increments. The purpose of this paper is the detailed examination, both theoretical and empirical, of a model in which both assumptions are removed. This model was introduced as the Markov tree (MT) model in our earlier work (Bhat and Kumar, 2010). The name of the model indicates that the tree is a generalization of the standard binomial tree, where the up/down factors at step $n+1$ depend on the direction of the step taken at step n. This is illustrated in Fig. 1. Though the description of the model is simple, and though it contains only two additional static parameters (σ^+ and σ^-) that must be estimated from data, the MT model leads to a number of non-trivial properties with significant consequences for option pricing.

By construction, the MT model accounts for the serial dependence of log returns. As we show, the distribution generated by the MT model is very closely approximated by a mixture of normals. Though this topic is not pursued further here, the MT model is a tree model that could be used to price path-dependent options. Hence the MT model can be seen as combining the strengths of normal mixture models, non-IID models, and tree methods all within the framework of risk-neutral pricing. In this paper, we derive an accurate, computationally efficient, closed-form approximation to the MT model option price. We go on to subject our model to out-of-sample comparisons against market prices and Black-Scholes model prices.

*School of Natural Sciences, University of California, Merced, 5200 N. Lake Rd., Merced, CA, 95343 (USA).
†email: hbhat@ucmerced.edu

The MT model incorporates several features that have been studied separately in the literature. The first such feature is the use of a mixture of normals. It is widely accepted that the observed distribution of daily log returns for stocks has heavier tails than the normal distribution, skewness, and positive excess kurtosis (Cont, 2001; Campbell et al., 1997; Barone-Adesi, 1985; Longin, 2005; Behr and Pötter, 2009). Many distributions have been proposed to match these properties. These distributions can be classified into parametric and non-parametric models—for an extensive list, see (Jackwerth, 1999). Parametric models include generalized distributions (Eberlein and Keller, 1995) and mixture distributions (Kon, 1984). Empirical tests (Behr and Pötter, 2009) conclude that normal mixture models fit observed log returns better than other generalized parametric models. In recent work, mixture distributions have been used in both option pricing and portfolio optimization (Tan and Chu, 2012; Cai and Kou, 2011; Ramponi, 2011; Buckley et al., 2008; Brigo and Mercurio, 2002; Ritchey, 1990) with success.

The second feature of the MT model is the non-IID process used to model the underlying asset dynamics. The study of (Niederhoffer and Osborne, 1966) was one of the first to examine serial dependence of log returns, providing strong evidence of dependence in tick differences. Daily returns have been studied by many authors, e.g., (Fielitz and Bhargava, 1973; Fielitz, 1975; Ding et al., 1993; Taylor, 2007), providing considerable evidence that daily returns are not independent. For returns sampled at longer intervals, i.e., monthly or yearly, the evidence is inconclusive (Sewell, 2011). Note that the short-term dependence of log returns need not invalidate the weak form of the efficient market hypothesis (Fama, 1970).

Several option pricing models have been proposed that allow for serial dependence of the underlying asset's returns. A direct approach is to explicitly account for dependence on the past in the underlying asset model. This strategy has been pursued with Markov and semi-Markov processes (Janssen et al., 1997; D'Amico et al., 2009), jump-diffusion processes with non-IID jumps (Camara and Li, 2008), and stochastic delay differential equations (SDDEs) (Appleby et al., 2012a,b; Chang et al., 2011, 2010; Swords and Appleby, 2010; Wu M. et al., 2008; Chang and Youree, 2007; Kazmerchuk et al., 2007; Arriojas et al., 2007). In the case of SDDE models, obtaining a closed-form approximation for the option price is much more difficult than for the MT model. Furthermore, when SDDE models are proposed in the literature, the performance of the models has not been tested using market data.

Another approach that yields a non-IID model is to introduce the concept of a regime; in a regime-switching model, a stochastic process (typically, a Markov chain) drives the regime from one state to another, and model parameters such as volatility and the risk-free rate are functions of the regime state (Mamon and Rodrigo, 2005; Aingworth et al., 2006; Basu and Ghosh, 2009). Finally, we note that in the framework of stochastic and/or GARCH volatility (Heston, 1993; Heston and Nandi, 2000) models, non-IID returns are a side effect of a volatility process that allows for memory.

The general outline of this paper is as follows. In Section 2, we review the definition of the MT model and the properties established in (Bhat and Kumar, 2010). Then, in Section 3, we prove that the tree is recombinant and give an exact formula for the option price. The exact formula relies on a discrete p.m.f. (probability mass function) that becomes prohibitively difficult to compute as the size of the time step vanishes. Therefore, in Section 4, we approximate the p.m.f. by a continuous p.d.f. (probability density function), which turns out to be a mixture of normal distributions. In Section 5, we use the approximate continuous p.d.f. to derive a closed-form option price. In Section 6, we conduct out-of-sample empirical tests that show that the MT model's prices are very close to market prices. In the same section, we give our conclusions and directions for further research.

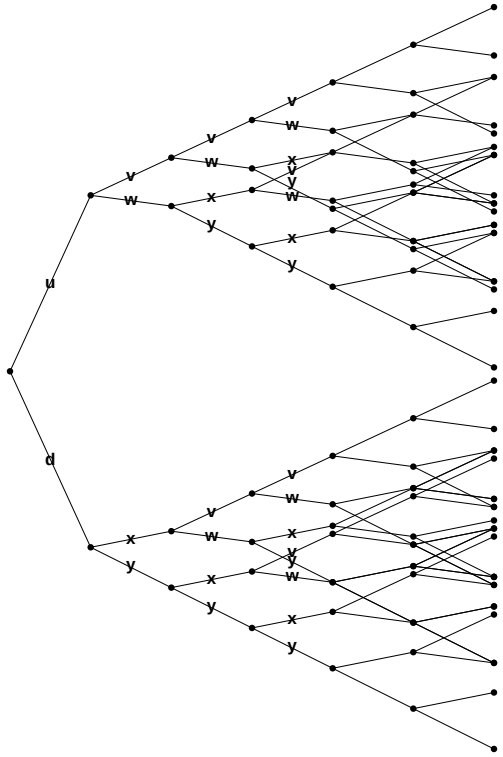

Figure 1: Tree of depth $n = 6$ showing recombination of paths of the underlying asset in the MT model. The asset begins with price S_0 and is multiplied by the weights along the path. For example, a possible path of length 3 shown here is S_0uwx. Both the probabilities and the outcomes of S_{n+1}/S_n depend on whether S_n/S_{n-1} was an upward or downward movement. In this way, the tree accounts for first-order Markov dependence of log returns. At depth n, there are $n^2 - n + 2$ possible states, as shown in Section 3.2.

2 Review of the Markov Tree Model

To keep this paper self-contained, we review the main points of (Bhat and Kumar, 2010).

2.1 Underlying asset dynamics

The Markov tree (MT) models the up and down movements in the underlying asset as a first-order Markov chain. The total time to expiration Y (in years) is divided into N equispaced steps with $\Delta t = Y/N$ years. Let S_n be the asset price at time step n. For $n = 0$, let the magnitudes of the up/down movements be u and d, and let the probability of an up movement be q. Then $P(S_1 = S_0u) = q$, and $P(S_1 = S_0d) = 1 - q$.

For $n \geq 1$, the model departs from the standard binomial option pricing model. We define two new events $S_n^+ = \{S_n \geq S_{n-1}\}$ and $S_n^- = \{S_n < S_{n-1}\}$; these events correspond to an upward (S_n^+) or downward (S_n^-) movement in price from time step $n - 1$ to time step n.

For $n \geq 1$, we define the evolution of the tree using two new probabilities q^+ and q^-:

$$P(S_{n+1} = S_n v | S_n^+) = q^+, \quad P(S_{n+1} = S_n w | S_n^+) = 1 - q^+, \quad \text{and}$$
$$P(S_{n+1} = S_n x | S_n^-) = q^-, \quad P(S_{n+1} = S_n y | S_n^-) = 1 - q^-.$$

We have introduced four symbols v, w, x and y that are factors by which the stock price at any step is allowed to change. The stock price goes up or down by a factor of v or w (respectively, x or y) if the stock price increased (respectively, decreased) in the previous step of the tree. See Fig. 1 for a pictorial representation of the tree, showing its recombinant behavior for $n \geq 4$. In Section 3, we show that the number of nodes at depth n of the tree is $n^2 - n + 2$, far less than the worst-case behavior of 2^n for a non-recombinant tree.

2.2 Martingale property

Let r be the risk-free rate of interest. We define

$$q = \frac{\exp(r\Delta t) - d}{u - d}, \qquad q^+ = \frac{\exp(r\Delta t) - w}{v - w}, \qquad q^- = \frac{\exp(r\Delta t) - y}{x - y}. \tag{1}$$

Then one checks that $E[S_1|S_0] = uS_0q + dS_0(1-q) = e^{r\Delta t}S_0$, and for $n \geq 1$,

$$\begin{aligned}
E[S_{n+1}|S_n, \ldots, S_0] &= E[S_{n+1}|S_n, \ldots, S_0, S_n^+]P(S_n^+) + E[S_{n+1}|S_n, \ldots, S_0, S_n^-]P(S_n^-) \\
&= \left[vS_nq^+ + wS_n(1-q^+)\right] P(S_n^+) + \left[xS_nq^- + yS_n(1-q^-)\right] P(S_n^-) \\
&= e^{r\Delta t}S_nP(S_n^+) + e^{r\Delta t}S_nP(S_n^-) \\
&= e^{r\Delta t}S_n.
\end{aligned}$$

This implies that the discounted process $\overline{S}_n = e^{-rn\Delta t}S_n$ is a martingale under the risk-neutral probabilities (1). By standard arguments, this implies that the MT model does not admit arbitrage (see (Shreve, 2004, Chapter 2.4)).

2.3 Risk-neutral pricing

Let us now use the Markov tree to price a European call option with strike K and spot price S_0. We defer all details on statistical estimation of the parameters r, u, d, v, w, x, and y to Section 6—at the moment, we take these parameters as given. Then, using the Markov tree and (1), we generate a risk-neutral p.m.f. for the random variable S_N, the share price of the underlying asset at the time of expiry. Let J_N denote the set of possible outcomes of S_N. Finally, for a random variable $X(\omega)$, let $X_+(\omega)$ denote the random variable

$$X_+(\omega) = \begin{cases} X(\omega) & X(\omega) > 0 \\ 0 & X(\omega) \leq 0. \end{cases}$$

Then the MT model's call options price is the discounted expected value of the option's payoff:

$$C = e^{-rY}E[(S_N - K)_+] \tag{2}$$

Using the exact p.m.f. generated by the Markov tree, we have

$$C = e^{-rY} \sum_{\substack{\sigma \in J_N \\ \sigma > K}} (\sigma - K) P(S_N = \sigma). \tag{3}$$

The MT model has been subjected to out-of-sample tests against market prices of European options. We used the MT model to price options on six stocks from the French CAC-40 index. Across 44 days of testing, the MT model's prices more closely matched market prices than the Black-Scholes model's prices (Bhat and Kumar, 2010).

3 Markov Tree Generation and Computational Tractability

Here we establish that the maximum number of possible states in a Markov tree of depth n is $n^2 - n + 2$. We also give a method for computing the p.m.f. of S_n, the underlying asset price after n steps of the tree.

3.1 Persistent random walk

The time evolution of $\tilde{S}_n = \log S_n$ under the Markov tree is equivalent to a persistent random walk on the real line, where both the size and direction of the walker's step at time step $n+1$ depends on the direction of the step taken at time step n:

$$\tilde{S}_{n+1}(\omega) = \tilde{S}_n(\omega) + G(\mathcal{H}(\tilde{S}_n - \tilde{S}_{n-1}), \omega), \tag{4}$$

where $\mathcal{H}$ is the Heaviside function

$$\mathcal{H}(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0, \end{cases}$$

and $G(1,\omega)$, $G(0,\omega)$ are random variables with p.m.f.'s

$$\begin{aligned} P(G(1,\omega) = \log v) &= q^+, \quad P(G(1,\omega) = \log w) = 1 - q^+, \quad \text{and} \\ P(G(0,\omega) = \log x) &= q^-, \quad P(G(0,\omega) = \log y) = 1 - q^- \end{aligned}$$

for $n \geq 2$. For $n = 1$, the p.m.f. of $G(1,\omega)$ and $G(0,\omega)$ is given by

$$\begin{aligned} P(G(1,\omega) = \log u) &= q \\ P(G(1,\omega) = \log d) &= 1 - q. \end{aligned}$$

We assume $\log u$, $\log d$, $\log v$, $\log w$, $\log x$, $\log y$ are all non-zero, so that $P(\tilde{S}_n = \tilde{S}_{n-1}) = 0$.

3.2 Number of states in a tree of fixed depth

For the moment, we ignore the size of the walker's steps and focus only on their direction. If the walker moves to the right (respectively, left), we call that heads H (respectively, tails T). The walk after n steps can be regarded as a random sequence of heads H and tails T.

Let n_H (respectively, n_T) be 1 if the first element is H (respectively, T) and 0 otherwise. Let n_{HH}, n_{HT}, n_{TH},

and n_{TT} denote the number of subsequences of the form HH, HT, TH, and TT. Then

$$n_{HH} + n_{HT} + n_{TH} + n_{TT} = n - 1. \tag{5}$$

Let $\mathbf{v} = (n_H, n_T, n_{HH}, n_{HT}, n_{TH}, n_{TT})$. The final position of the walker is $\tilde{S}_n = \tilde{S}_0 + \mathbf{s} \cdot \mathbf{v}$ where $\mathbf{s} = (\log u, \log d, \log v, \log w, \log x, \log y)$. Hence enumerating all possible vectors $\mathbf{v}$ is equivalent to enumerating all possible outcomes of $\tilde{S}_n$.

Suppose that the sequence starts with H. Let t denote the number of *transitions*:

$$t = n_{HT} + n_{TH}. \tag{6}$$

Now t can be anything from 0 to $n-1$. Given t, we know n_{HT} and n_{TH}, since transitions must alternate H to T and T to H. For $t = 0$, there is only one sequence $HHH \cdots H$.

For $t = 1, 2, \ldots, n-1$, a walk with t transitions is a sequence of $t+1$ blocks, with odd blocks consisting of consecutive H's and even blocks consisting of consecutive T's. We start with the sequence $HTHTHT \cdots$ of length $t+1$. To convert this into a walk of length n, we must insert extra H's into the H blocks and extra T's into the T blocks, inserting $n-t-1$ elements in total. Now n_{HH} is the number of H's inserted, so it can be anything from 0 to $n-t-1$, for $n-t$ possibilities in total. Once we know n_{HH}, we solve for n_{TT} using (5).

For a walk of length n starting with H, the number of possible $\mathbf{v}$'s is $1 + \sum_{t=1}^{n-1}(n-t) = 1 + \dfrac{n(n-1)}{2}$. Twice this number is $n^2 - n + 2$, the total number of possibilities for $\mathbf{v}$. Note that the regime-switching model of (Aingworth et al., 2006), if used with two volatility states, results in a different tree that also has quadratic complexity.

3.3 Markov tree probability mass function

Now let us assume $\mathbf{v}$ is given and count how many walks correspond to that same $\mathbf{v}$. Starting with H, there are $a = n_{TH} + 1$ blocks of heads and $b = n_{HT}$ blocks of tails.

Given n_{HH} and n_{TT}, to obtain the walk we must decide how many of the n_{HH} extra heads to insert into each block, with the total being n_{HH}. The number of such possibilities is the number of weak compositions of n_{HH} into a nonnegative integers, $\binom{n_{HH}+a-1}{a-1}$.

We must also decide how many of the n_{TT} extra tails to insert into each block, with the total being n_{TT}. The number of such possibilities is the number of weak compositions of n_{TT} in b nonnegative integers, $\binom{n_{TT}+b-1}{b-1}$.

Hence the number of walks that start with H and correspond to $\mathbf{v}$ is

$$\#(\mathbf{v}) = \binom{n_{HH}+a-1}{a-1}\binom{n_{TT}+b-1}{b-1} = \binom{n_{HH}+n_{TH}}{n_{TH}}\binom{n_{TT}+n_{HT}-1}{n_{HT}-1}. \tag{7}$$

If instead the walk starts with T, the only difference is that $a = n_{TH}$ and $b = n_{HT} + 1$ and we obtain

$$\#(\mathbf{v}) = \binom{n_{HH}+a-1}{a-1}\binom{n_{TT}+b-1}{b-1} = \binom{n_{HH}+n_{TH}-1}{n_{TH}-1}\binom{n_{TT}+n_{HT}}{n_{HT}} \tag{8}$$

as the number of walks.

Once we know how many ways there are of reaching $\tilde{S}_n$ from $\tilde{S}_0$, we can compute

$$P(\tilde{S}_n = \tilde{S}_0 + \mathbf{s} \cdot \mathbf{v}) = \#(\mathbf{v})\, \mathbf{q}^{\mathbf{v}}, \tag{9}$$

where $\mathbf{q} = (q, 1-q, q^+, 1-q^+, q^-, 1-q^-)$ and $\mathbf{q}^{\mathbf{v}} = \prod_{j=1}^{6} {q_j}^{v_j}$. In this way, the entire p.m.f. of $\tilde{S}_n$ is determined.

Care must be used when applying the above formulas, as they do not detect whether the walk is allowed or not. If the walk corresponding to $\mathbf{v}$ is allowed, then the above formulas give the number of walks.

This begs the question of enumerating all allowed $\mathbf{v}$'s at a fixed depth n. This can be done using the following algorithm, which works for all walks that start with H (so that $n_H = 1$):

print $\mathbf{v} = (1, 0, n-1, 0, 0, 0)$
for $t = 1 \to n-1$ **do**
  $n_{HT} = \lceil t/2 \rceil$
  $n_{TH} = \lfloor t/2 \rfloor$
  for $n_{HH} = 0 \to n-t-1$ **do**
    $n_{TT} = (n-t-1) - n_{HH}$
    print $\mathbf{v} = (1, n_{HH}, n_{HT}, n_{TH}, n_{TT})$
  end for
end for

To enumerate all walks that start with T (so that $n_T = 1$), we use the same algorithm as above with two minor changes: (i) switch the definitions of n_{HT} and n_{TH}; (ii) change the $t = 0$ output of $\mathbf{v}$ to be $\mathbf{v} = (0, 1, 0, 0, 0, n-1)$. Using both algorithms, we produce a list of all allowed $\mathbf{v}$'s at a fixed depth n.

4 Continuous Approximation of the Markov Tree

We can see from (2) that the key ingredient in computing the Markov tree options price is taking the expected value of the payoff function with respect to the p.m.f. (9) generated by the tree. Though we have developed an efficient algorithm to generate all states of the tree, the quantity $\#(\mathbf{v})$ defined by (7) and (8) is difficult to compute in finite-precision arithmetic due to the large binomial coefficients involved. In this section, we develop a closed-form continuous p.d.f. that closely approximates the discrete Markov tree p.m.f.

The p.d.f., which turns out to be a mixture of normals, also yields an intuitive understanding of the distribution of asset prices generated by the Markov tree. This understanding will lead us to a reasonable method to statistically estimate the parameters u, v, and x from market data.

4.1 Recursion

To develop a continuous approximation, we first rewrite the discrete-time process (4) as a recursion. We assume all movements are symmetric about one (i.e., $d = 1/u$, $w = 1/v$, $y = 1/x$) and define

$$l_u = \log u = -\log d \tag{10a}$$

$$l_1 = \log v = -\log w \tag{10b}$$

$$l_2 = \log x = -\log y \tag{10c}$$

We assume l_u, l_1, and l_2 are all positive.

Let $R(n, \tilde{s})$ be the probability of reaching a value $\tilde{s}$ on the real line in n steps by moving to the *right* (in the positive direction on $\mathbb{R}$) in the n-th step. Similarly, let $L(n, \tilde{s})$ be the probability of reaching the value $\tilde{s}$ in n steps by moving to the *left* (in the negative direction on $\mathbb{R}$) in the n-th step.

In the Markov tree, since $\log v$ and $\log x$ are the only positive steps allowed, $R(n, \tilde{s})$ is the probability of reaching $\tilde{s}$ in n steps by taking either a $\log v$ step or a $\log x$ step in the n-th step. If the n-th step was a $\log v$ step, then after $n-1$ steps, the walker was at $\tilde{s} - l_1$ and had reached there by taking the $(n-1)$-th step to the right. The probability of the walker reaching this position in this way after $n-1$ steps is $R(n-1, \tilde{s} - l_1)$. Similarly, if the n-th step was a $\log x$ step, then after $n-1$ steps, the walker was at $\tilde{s} - l_2$ and had reached there by taking the $(n-1)$-th step to the left. The probability of the walker reaching this position in this way after $n-1$ steps is $L(n-1, \tilde{s} - l_2)$.

Putting things together, we obtain

$$R(n, \tilde{s}) = q^+ R(n-1, \tilde{s} - l_1) + q^- L(n-1, \tilde{s} - l_2). \tag{11}$$

Next, since $\log w$ and $\log y$ are the only negative steps in the Markov tree, $L(n, \tilde{s})$ is the probability of reaching $\tilde{s}$ in n steps by taking a $\log w$ step or a $\log y$ step in the n-th step. If the n-th step was a $\log w$ step, then the walker was at $\tilde{s} + l_1$ after $n-1$ steps and had reached there by taking the $(n-1)$-th step to the right. The probability of the walker reaching this position in this way after $n-1$ steps is $R(n-1, \tilde{s} + l_1)$. Similarly, if the n-th step was a $\log y$ step, then the random walker was at $\tilde{s} + l_2$ after $n-1$ steps and had reached there by taking the $(n-1)$-th step to the left. The probability of the walker reaching this position in this way after $n-1$ steps is $L(n-1, \tilde{s} + l_2)$.

Putting things together, we obtain

$$L(n, \tilde{s}) = (1 - q^+) R(n-1, \tilde{s} + l_1) + (1 - q^-) L(n-1, \tilde{s} + l_2). \tag{12}$$

4.2 Exact solution in Fourier space

We introduce the following forward and inverse Fourier transform pair, with the variable k as the Fourier conjugate variable to $\tilde{s}$:

$$\hat{f}(k) = \int_{\mathbb{R}} f(\tilde{s}) e^{-ik\tilde{s}} \, d\tilde{s}, \qquad f(\tilde{s}) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(k) e^{ik\tilde{s}} \, dk. \tag{13}$$

Define

$$M = \begin{bmatrix} q^+ e^{-ikl_1} & q^- e^{-ikl_2} \\ (1-q^+) e^{ikl_1} & (1-q^-) e^{ikl_2} \end{bmatrix}. \tag{14}$$

Then, taking the Fourier transforms of both sides of (11) and (12), we are able to put the system into matrix-vector form and solve:

$$\begin{bmatrix} \hat{R}(n,k) \\ \hat{L}(n,k) \end{bmatrix} = M \begin{bmatrix} \hat{R}(n-1,k) \\ \hat{L}(n-1,k) \end{bmatrix} = M^{n-1} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix}. \tag{15}$$

Let $P(n, \tilde{s}) = R(n, \tilde{s}) + L(n, \tilde{s})$. Then $P(n, \tilde{s})$ is the p.d.f. of the random variable $\tilde{S}_n$. The Fourier transform of the p.d.f. is given by $\hat{P}(n,k) = \hat{R}(n,k) + \hat{L}(n,k)$. We compute $\hat{P}(n,k)$ by left multiplying equation (15) with the row vector $\mathbf{1}^t$:

$$\hat{P}(n,k) = \mathbf{1}^t M^{n-1} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix}. \tag{16}$$

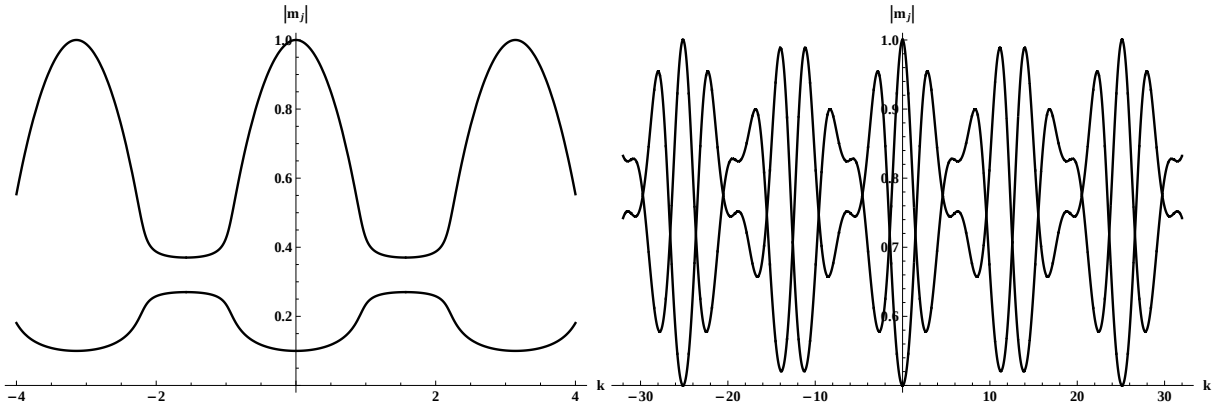

Figure 2: Moduli of the eigenvalues m_1, m_2 of the matrix M defined in (14). We plot $|m_j|$ as a function of Fourier variable k to show that, for almost all values of k, the eigenvalues are in the interior of the unit disc in $\mathbb{C}$. For the plot on the left, we set $l_1 = l_2 = 1$, $q^+ = 3/5$, $q^- = 1/2$. For the plot on the right, we set $l_1 = 5/4$, $l_2 = 1$, $q^+ = 1/5$, $q^- = 7/10$. We obtain similar behavior for many other parameter choices.

Since M is diagonalizable, raising it to the n-th power is computationally economical and we can easily compute $\hat{P}(n,k)$. By construction of the Markov tree, $R(1,\tilde{s}) = q\delta\left(\tilde{s} - (\tilde{S}_0 + l_u)\right)$ and $L(1,\tilde{s}) = (1-q)\delta\left(\tilde{s} - (\tilde{S}_0 - l_u)\right)$, where δ is a point mass (Dirac delta).

4.3 Numerical solution in real space

In the numerical inversion of (16), the only difficulty that might possibly arise would be that the spectrum of M lies too close to the unit circle in $\mathbb{C}$. For this reason, we numerically explore the spectrum of M in Fig. 2. Let m_1 and m_2 be the eigenvalues of M. We plot the moduli $|m_1|$ and $|m_2|$ as functions of k for two different sets of parameters. The two plots shown there are representative; the spectrum of M is well-behaved.

To invert the Fourier transform (16) and obtain the p.d.f. of $\tilde{S}_n$, we use the algorithm described by (Inverarity, 2003). This approach to finding the p.d.f. is faster and more accurate than Taylor expanding the right hand side of equations (11) and (12) about l_1 and l_2 and then numerically solving the partial differential equation thus obtained.

Note that even though numerical Fourier inversion of (16) yields a fast, accurate approximation to the p.d.f. of $\tilde{S}_n$, the method has two deficiencies that prevent us from using it to price options: (i) it does not yield an analytical expression for the p.d.f., and (ii) it does not provide any intuition on how to statistically estimate the parameters u, v, and x. We will therefore use the p.d.f. obtained by numerical inversion of (16) only to compare against the true Markov tree p.m.f. (9) and the asymptotic approximation that we derive next.

4.4 Asymptotic solution in real space

We now use generating functions to derive an asymptotic approximation to the p.d.f. (Rudnick and Gaspari, 2004, Chapter 5.2). For $z \in \mathbb{C}$, we define

$$\rho(z,k) = \sum_{n=0}^{\infty} \hat{R}(n+1,k) z^n$$
$$\lambda(z,k) = \sum_{n=0}^{\infty} \hat{L}(n+1,k) z^n.$$

The functions ρ and λ are generating functions for R and L, respectively. Using (15), we can write

$$\begin{aligned}
\begin{bmatrix} \rho(z,k) \\ \lambda(z,k) \end{bmatrix} &= \sum_{n=0}^{\infty} M^n z^n \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix} \\
&= (\mathrm{I} - Mz)^{-1} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix} \\
&= \frac{1}{(1-zm_1)(1-zm_2)} \begin{bmatrix} 1-(1-q^-)e^{ikl_2}z & q^- e^{-ikl_2} z \\ (1-q^+)e^{ikl_1}z & 1-q^+e^{-ikl_1}z \end{bmatrix} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix}. \qquad (17)
\end{aligned}$$

Let p be the generating function for $\hat{P}$. Then

$$\begin{aligned}
p(z,k) &= \sum_{n=0}^{\infty} \hat{P}(n+1,k) z^n \\
&= \sum_{n=0}^{\infty} \left(\hat{R}(n+1,k) + \hat{L}(n+1,k) \right) z^n \\
&= \mathbf{1}^t \begin{bmatrix} \rho(z,k) \\ \lambda(z,k) \end{bmatrix}. \qquad (18)
\end{aligned}$$

Substituting (17) in (18) and carrying out the algebra, we have

$$p(z,k) = \frac{\hat{P}(1,k)}{(1-zm_1)(1-zm_2)} + z\frac{\gamma}{(1-zm_1)(1-zm_2)},$$

where

$$\gamma = \hat{R}(1,k)\left((1-q^+)e^{il_1k} - (1-q^-)e^{il_2k}\right) + \hat{L}(1,k)\left(q^- e^{-il_2k} - q^+ e^{-il_1k}\right),$$

independent of z. Continuing with the calculation, we get

$$\begin{aligned}
p(z,k) &= \frac{\hat{P}(1,k)}{(1-zm_1)(1-zm_2)} + z\frac{\gamma}{(1-zm_1)(1-zm_2)} \\
&= \frac{\hat{P}(1,k)}{m_1-m_2}\left(\frac{m_1}{1-zm_1} - \frac{m_2}{1-zm_2}\right) + z\frac{\gamma}{m_1-m_2}\left(\frac{m_1}{1-zm_1} - \frac{m_2}{1-zm_2}\right) \\
&= \frac{\hat{P}(1,k)}{m_1-m_2}\left(m_1\sum_{n=0}^{\infty} m_1^n z^n - m_2\sum_{n=0}^{\infty} m_2^n z^n\right) + z\frac{\gamma}{m_1-m_2}\left(m_1\sum_{n=0}^{\infty} m_1^n z^n - m_2\sum_{n=0}^{\infty} m_2^n z^n\right) \\
&= \frac{1}{m_1-m_2}\sum_{n=0}^{\infty}\left(\hat{P}(1,k)(m_1^{n+1}-m_2^{n+1}) + \gamma(m_1^n - m_2^n)\right)z^n. \qquad (19)
\end{aligned}$$

By definition, $\hat{P}(n+1,k)$ is given by the coefficient of z^n in the expansion of $p(z,k)$. The quantities m_1, m_2 and $\hat{P}(1,k)$ are all independent of z. Thus $\hat{P}(n+1,k)$ can simply be read off from (19):

$$\hat{P}(n+1,k) = \frac{1}{m_1-m_2}\left(\hat{P}(1,k)(m_1^{n+1}-m_2^{n+1}) + \gamma(m_1^n - m_2^n)\right). \qquad (20)$$

The above quantity represents the Fourier transform of the probability of reaching the value $\tilde{s}$ in n steps and matches the right-hand side of (16).

Since m_1, m_2, and γ all depend on k, we cannot expect to find a closed-form inverse Fourier transform of (20). However, the tail behavior of $P(n+1,\tilde{s})$ as $n\to\infty$ and $\tilde{s}\to\infty$ can be determined to a close approximation. To do this, we expand $\hat{P}(n+1,k)$ about $k=0$ and calculate the inverse Fourier transform of the leading terms. The leading terms represent the behavior in the tail where the higher spatial derivatives of $P(n+1,\tilde{s})$ are nearly zero. For justification of this procedure, we refer to (Lighthill, 1958).

Let m_1 (respectively, m_2) be the eigenvalue of M with larger (respectively, smaller) modulus. We expand these eigenvalues in powers of k:

$$m_1 = 1 + i\zeta_{11}k - \zeta_{12}k^2 + O(k^3) \qquad (21)$$

$$m_2 = (q^+ - q^-) + i\zeta_{21}k - \zeta_{22}k^2 + O(k^3) \qquad (22)$$

The expressions for the ζ_{lm} coefficients are lengthy and shall be omitted. The first step of the Markov tree gives

$$\begin{aligned}
\hat{P}(1,k) &= \hat{R}(1,k) + \hat{L}(1.k) \\
&= qe^{-i(\tilde{S}_0+l_u)k} + (1-q)e^{-i(\tilde{S}_0-l_u)k}.
\end{aligned}$$

Define the constants

$$\begin{aligned}
F_1 &= e^{-i(\tilde{S}_0+l_u)k}, \quad F_2 = e^{-i(\tilde{S}_0-l_u)k}, \\
\alpha &= (1-q^+)e^{il_1k} - (1-q^-)e^{il_2k}, \quad \beta = q^-e^{-il_2k} - q^+e^{-il_1k}.
\end{aligned}$$

We express $\gamma = qF_1\alpha + (1-q)F_2\beta$ and $\hat{P}(1,k) = qF_1 + (1-q)F_2$. Since $|m_1| > |m_2|$ and $|m_{1,2}| \le 1$, when n is

large, we get

$$\begin{aligned}\hat{P}(n+1,k) &\sim \frac{1}{m_1-m_2}\left(\hat{P}(1,k)m_1^{n+1}+\gamma m_1^n\right)\\ &= \frac{1}{m_1-m_2}\left(qF_1(m_1^{n+1}+\alpha m_1^n)+(1-q)F_2(m_1^{n+1}+\beta m_1^n)\right)\\ &\sim m_1^{n-1}\left(qF_1(m_1+\alpha)+(1-q)F_2(m_1+\beta)\right). \qquad (23)\end{aligned}$$

Let us concentrate on the first term and approximate it to $O(k^3)$. We get

$$\begin{aligned}m_1^{n-1}qF_1(m_1+\alpha) &= \exp\left(\log q+\log(F_1m_1+F_1\alpha)+(n-1)\log m_1\right)\\ &= q\exp\left(\log(F_1m_1+F_1\alpha)+(n-1)\log m_1\right) \qquad (24)\\ &\sim q\exp\left(-\mu_1 ik-\frac{\sigma_1^2}{2}k^2\right). \qquad (25)\end{aligned}$$

To pass from (24) to (25), we expand the argument of the exponential function in powers of k. Note that F_1, m_1, and α are all functions of k defined above. The real coefficients μ_1 and σ_1 are defined in detail in the Appendix. Proceeding analogously for the second term in (23), we get

$$m_1^{n-1}(1-q)F_2(m_1+\beta)\sim(1-q)\exp\left(-\mu_2 ik-\frac{\sigma_2^2}{2}k^2\right),$$

where μ_2 and σ_2 are real constants defined in the Appendix. We now express $\hat{P}(n,k)$ as

$$\hat{P}(n+1,k)\sim q\exp\left(-\mu_1 ik-\frac{\sigma_1^2}{2}k^2\right)+(1-q)\exp\left(-\mu_2 ik-\frac{\sigma_2^2}{2}k^2\right). \qquad (26)$$

Taking the inverse Fourier transform of both sides of (26), we obtain the approximate p.d.f.

$$P(n+1,\tilde{s})\sim f_{\tilde{s}}(\tilde{s},n+1):=\frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{(\tilde{s}-\mu_1)^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(\tilde{s}-\mu_2)^2}{2\sigma_2^2}\right). \qquad (27)$$

This shows that the p.d.f. of $\tilde{S}_{n+1}$ is well-approximated by a weighted mixture of two normals. The first normal $\mathcal{N}(\mu_1,\sigma_1^2)$ has weight q and the second normal $\mathcal{N}(\mu_2,\sigma_2^2)$ has weight $1-q$. Let the p.d.f. of the first (respectively, second) normal be g_1 (respectively, g_2), so that we can write

$$f_{\tilde{s}}(\tilde{s},n+1)=qg_1(\tilde{s},n+1)+(1-q)g_2(\tilde{s},n+1). \qquad (28)$$

4.5 Comparison of the distribution functions for the Markov tree

We now have two continuous densities to compare against the Markov tree p.m.f. To enable a fair comparison between discrete and continuous random variables, we compare cumulative distribution functions (c.d.f.'s). In Fig. 3, we plot the c.d.f.'s obtained from the following probability mass/density functions: MT, the exact Markov tree p.m.f. (9), FT, the p.d.f. obtained by numerical inversion of the Fourier transform (16), and Asym, the p.d.f. (27) obtained by asymptotic approximation. Table 1 shows the parameters used in the comparison in each of the panels.

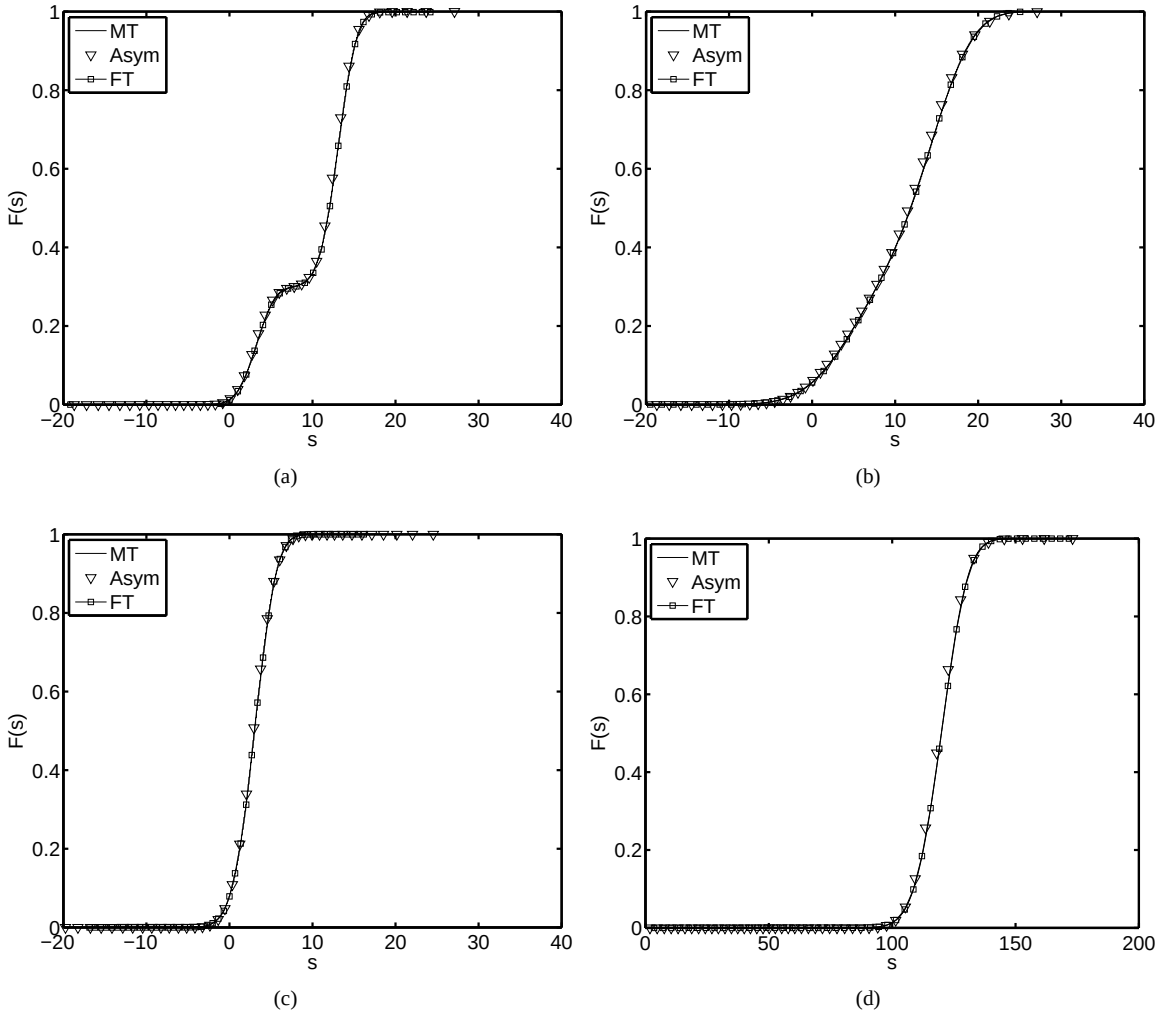

Figure 3: Comparison of cumulative distribution functions for MT, the exact Markov tree p.m.f. (9), FT, the p.d.f. obtained by numerical inversion of the Fourier transform (16), and Asym, the p.d.f. (27) obtained by asymptotic approximation. Table 1 shows the parameters used in the comparison in each of the panels.

Panel of Fig. 3	l_u	l_1	l_2	q	q^+	q^-	N	$\|\mathrm{FT}-\mathrm{MT}\|_\infty$	$\|\mathrm{Asym}-\mathrm{MT}\|_\infty$
3a	5.0	0.2	0.3	0.7	0.4	0.8	150	0.0097	0.0362
3b	5.0	0.2	0.3	0.7	0.8	0.4	150	0.0042	0.0247
3c	0.05	0.2	0.3	0.5	0.3	0.7	150	0.0117	0.0320
3d	0.05	0.4	0.6	0.5	0.8	0.7	500	0.0065	0.0403

Table 1: Details of parameters used for each panel in Fig. 3 and numerical values of the errors.

We see that the FT and Asym c.d.f's closely approximate the exact MT c.d.f. There is nothing special about the parameter values chosen for the tests whose results are shown—for other parameter values, the approximations are just as good.

Table 1 also shows the error in the $\|\cdot\|_\infty$ norm for the FT and Asym approximations. The FT approximation is better than the Asym approximation; however, the deficiencies of the FT approximation noted at the end of Section 4.3 still apply.

Note that we have also conducted tests where we have compared the prices of European call options computed using the MT distribution against those computed using the Asym distribution. The differences are negligible. In what follows, we use the asymptotic normal mixture distribution (27) and (28) to price options.

5 Option Price

Pricing a European call option using the normal mixture distribution (27) and (28) is straightforward. Suppose Y is the time to expiration (in years) and S_Y is the random variable representing the spot price of the underlying asset at time of expiry. In this section, we will take $f_{\tilde{s}}(\tilde{s}, n+1)$ to be the p.d.f. of $\tilde{S}_Y$—in other words, we ignore the fact that this p.d.f. is only an approximation.

We recall (2) and evaluate the expected value using the p.d.f. (27):

$$C = e^{-rY} \int_K^\infty (s-K) f_s(s, Y)\, ds, \tag{29}$$

where r is the risk-free rate, K is the strike price, and $f_s(s, Y)$ is the p.d.f. of S_Y. If dt is the the duration in years of each time step, then the total number of steps required in the Markov tree is $n+1 = Y/dt$. We chose dt small enough such that $N \gg 100$.

To relate the density of S_Y to the density of $\tilde{S}_Y$, we start with

$$P(S_Y \le s) = P(\tilde{S} \le \tilde{s}) = \int_{-\infty}^{\log s} f_{\tilde{s}}(\tilde{s}, n+1)\, d\tilde{s},$$

where $\tilde{s} = \log s$. Taking derivatives of both sides with respect to s, we see that

$$f_s(s, Y) = \frac{1}{s} f_{\tilde{s}}(\tilde{s}, n+1).$$

Now we can continue the calculation from (29) and use the decomposition (28)

$$\begin{aligned}
Ce^{rY} &= \int_K^\infty s\frac{1}{s} f_{\tilde{s}}(\tilde{s}, t)\, ds - \int_K^\infty K\frac{1}{s} f_{\tilde{s}}(\tilde{s}, t)\, ds \\
&= \int_K^\infty f_{\tilde{s}}(\tilde{s}, t)\, ds - K\int_K^\infty \frac{1}{s} f_{\tilde{s}}(\tilde{s}, t)\, ds \\
&= q\int_K^\infty g_1(\tilde{s}, t)\, ds + (1-q)\int_K^\infty g_2(\tilde{s}, t)\, ds - Kq\int_K^\infty \frac{1}{s} g_1(\tilde{s}, t)\, ds - K(1-q)\int_K^\infty \frac{1}{s} g_2(\tilde{s}, t)\, ds.
\end{aligned}$$

The value of the European call option can then be expressed in terms of μ_1, μ_2, σ_1 and σ_2 as

$$Ce^{rY} = q \exp\left(\frac{\sigma_1^2}{2} + \mu_1\right) \Phi(x_1) + (1-q) \exp\left(\frac{\sigma_2^2}{2} + \mu_2\right) \Phi(x_2) - qK\Phi(x_3) - (1-q)K\Phi(x_4), \tag{30}$$

where Φ is the distribution function of the standard normal, and

$$x_1 = \frac{\mu_1 - \sigma_1^2 - \log K}{\sigma_1}, \quad x_2 = \frac{\mu_2 - \sigma_2^2 - \log K}{\sigma_2}$$
$$x_3 = \frac{\mu_1 - \log K}{\sigma_1}, \quad x_4 = \frac{\mu_2 - \log K}{\sigma_2}.$$

Suppose that the underlying stock does not pay a dividend. Then C is also the value of the American call option on the stock (Bouchaud and Potters, 2003).

6 Empirical Results

In this section, we price options on 89 non-dividend-paying stocks from the S&P 500. Our goal is to compare Black-Scholes model prices and Markov Tree model prices against market prices. In what follows, we use a risk-free rate of interest $r = 0.01$, corresponding to the annualized rate of return for the shortest-term US Treasury bills during the time period of testing.

6.1 Parameter estimation

To price options using the MT model, we must statistically estimate three volatility parameters $(\sigma, \sigma^+, \sigma^-)$ from data. Assuming we have these parameters, we define

$$l_u = \sigma\sqrt{\Delta t}, \quad l_1 = \sigma^+\sqrt{\Delta t}, \quad l_2 = \sigma^-\sqrt{\Delta t}. \tag{31}$$

Then u, d, v, w, x, and y are defined by (10), enabling us to calculate the risk-neutral probabilities via (1), the mixture parameters (μ_j, σ_j) defined in the Appendix, and the call option price defined by (30).

For the Black-Scholes model, we need only estimate one volatility parameter σ. In our tests, we estimate σ using the sample annualized volatility $\hat{\sigma}$, the calculation of which proceeds via standard procedures described, for example, by (Hull, 2009). We use the same $\hat{\sigma}$ as our estimate for σ in the MT model.

We use two primary methods to estimate the volatility parameters $\sigma^{\pm}$:

1. *Naive Method.* We start with a time series of log returns: $Z = \{z_1, z_2, \ldots, z_\nu\}$, where $z_j = \log(S_j/S_{j-1})$ and S_j is the adjusted closing price for the stock on day j. We now form two disjoint subsets of Z:

$$Z^+ = \{z_j \in Z \mid z_{j-1} \geq 0\}, \qquad Z^- = \{z_j \in Z \mid z_{j-1} < 0\}$$

In words, Z^+ (respectively, Z^-) are the log returns on days for which the previous day's log return was non-negative (respectively, negative). We then compute

$$\hat{\sigma}^+ = \kappa \,\mathrm{mean}\left|Z^+ - \mathrm{mean}\left(Z^+\right)\right|, \qquad \hat{\sigma}^- = \kappa \,\mathrm{mean}\left|Z^- - \mathrm{mean}\left(Z^-\right)\right|. \tag{32}$$

Without the scaling factor κ, the quantity on the right-hand side is the mean absolute deviation of Z^+ or Z^-. The factor $\kappa = \sqrt{\pi/2}$ is included so that $\hat{\sigma}^\pm$ scales like the sample standard deviation (Kendall, 1944).

In this method, which is termed "MT naive" in the remainder of this paper, we use $\hat{\sigma}^\pm$ as our statistical estimates for $\sigma^\pm$. Note that past/present options prices are not used at all. The only market prices that are used are historical adjusted closing prices of the underlying stock. Hence our estimates $\hat{\sigma}^\pm$ do not depend on the strike price or time to expiry of the option that we are pricing.

2. *Regression Method.* In this method, we start with tables of end-of-day market prices of options. If we are interested in pricing options today, we look at yesterday's tables. We suppose there is one table for each stock symbol; each table lists a number of options with different strikes and expiration dates. Given $(\sigma, \sigma^+, \sigma^-)$ and all the parameters for the options in the table, we can use the MT model to generate a corresponding table of model prices.

   For each stock symbol, we use the algorithm of (Nelder and Mead, 1965) to search numerically for the optimal values (σ_*^+, σ_*^-) that minimize the error between the tables of market and model prices. Through this optimization, σ is set equal to the sample volatility $\hat{\sigma}$ described above. We also compute the estimates (32). In this way, we obtain for each stock symbol five values: $(\hat{\sigma}, \hat{\sigma}^+, \hat{\sigma}^-, \sigma_*^+, \sigma_*^-)$.

   Running the same procedure for all 89 stocks yields a matrix D of size 89×5. We treat each column of D as a vector with boldfaced labels $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^+, \hat{\boldsymbol{\sigma}}^-, \boldsymbol{\sigma}_*^+, \boldsymbol{\sigma}_*^-$. Our idea is to use the information contained in D to construct a model that uses one or more of the raw inputs $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^+, \hat{\boldsymbol{\sigma}}^-$ to predict the optimal values $\boldsymbol{\sigma}_*^+, \boldsymbol{\sigma}_*^-$. In what follows, we use $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ to denote residual errors.

   We first fit two ordinary least squares (OLS) linear regression models. In the first linear model, the response variables $\boldsymbol{\sigma}_*^\pm$ depend only on the raw volatility $\hat{\boldsymbol{\sigma}}$:

$$\boldsymbol{\sigma}_*^+ = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} \end{bmatrix} \boldsymbol{\eta}_\mathbf{1}^+ + \boldsymbol{\varepsilon}_\mathbf{1}^+ \tag{33a}$$

$$\boldsymbol{\sigma}_*^- = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} \end{bmatrix} \boldsymbol{\eta}_\mathbf{1}^- + \boldsymbol{\varepsilon}_\mathbf{1}^- \tag{33b}$$

   Since only one raw input is being used, we label the 2×1 vectors of regression coefficients by $\boldsymbol{\eta}_\mathbf{1}^\pm$. The adjusted R^2 values for this model can be found in the "One parameter" L^2 columns of Table 2.

   In the second linear model, the response variables $\boldsymbol{\sigma}_*^\pm$ depend on all three raw inputs $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^+, \hat{\boldsymbol{\sigma}}^-$; we now use $\boldsymbol{\eta}_\mathbf{3}^\pm$ to label the 4×1 vectors of regression coefficients:

$$\boldsymbol{\sigma}_*^+ = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} & \hat{\boldsymbol{\sigma}}^+ & \hat{\boldsymbol{\sigma}}^- \end{bmatrix} \boldsymbol{\eta}_\mathbf{3}^+ + \boldsymbol{\varepsilon}_\mathbf{3}^+ \tag{34a}$$

$$\boldsymbol{\sigma}_*^- = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} & \hat{\boldsymbol{\sigma}}^+ & \hat{\boldsymbol{\sigma}}^- \end{bmatrix} \boldsymbol{\eta}_\mathbf{3}^- + \boldsymbol{\varepsilon}_\mathbf{3}^- \tag{34b}$$

   The adjusted R^2 values for this model can be found in the "Three parameters" L^2 columns of Table 2.

   Comparing the adjusted R^2 values, we see that both linear models perform equally well. Both models fit fairly well for $\boldsymbol{\sigma}_*^+$, but the fit is poor for $\boldsymbol{\sigma}_*^-$, prompting explorations of nonlinear regression strategies.

   We report here the results of fitting two regression tree models (Breiman et al., 1984). In much the same way as we have done above, we first try a model that depends only on one raw input and then try a model that depends

on all three raw inputs. The first model can be written

$$\boldsymbol{\sigma}_*^+ = \psi_1^+(\hat{\boldsymbol{\sigma}}) + \boldsymbol{\delta}_\mathbf{1}^+ \tag{35a}$$

$$\boldsymbol{\sigma}_*^- = \psi_1^-(\hat{\boldsymbol{\sigma}}) + \boldsymbol{\delta}_\mathbf{1}^- \tag{35b}$$

The second model can be written

$$\boldsymbol{\sigma}_*^+ = \psi_3^+(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^+, \hat{\boldsymbol{\sigma}}^-) + \boldsymbol{\delta}_\mathbf{3}^+ \tag{36a}$$

$$\boldsymbol{\sigma}_*^- = \psi_3^-(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^+, \hat{\boldsymbol{\sigma}}^-) + \boldsymbol{\delta}_\mathbf{3}^- \tag{36b}$$

The adjusted R^2 values for models (35) and (36) can be found in Table 2, in the "Tree" columns with respective labels "One parameter" and "Three parameters." The fit for $\boldsymbol{\sigma}_*^-$ is much better for the tree models than it is for the L^2 models. Unlike the linear models, we also see that the model with more parameters fits better. Of course, we should keep in mind that these statements are made on the basis of in-sample performance. We conduct out-of-sample option pricing tests below.

In the remainder of this paper, the label "MT Reg" will be used to refer to the MT option pricing model where the parameters are estimated using the three-parameter tree regression model (36). Specifically, having trained the model ψ_3 using the previous day's option prices, we evaluate the model using today's raw estimates $\hat{\sigma}, \hat{\sigma}^+, \hat{\sigma}^-$. The outputs of the model, σ_*^+ and σ_*^-, are then used as statistical estimates of σ^+ and σ^-.

Note that the training of the tree regression model uses options data from the past. However, when we apply the tree regression model, we only need market data of stock log returns in order to compute the raw inputs $\hat{\sigma}$ and $\hat{\sigma}^\pm$. Just as in the naive model, the output of the tree regression model is therefore constant over the strikes and expiration dates of the options we will be pricing.

As a final note, we conjecture that there exists a more fundamental method for estimating the parameters $(\sigma, \sigma^+, \sigma^-)$. The regression approaches considered above should be viewed as attempts to infer the optimal model from data.

6.2 Empirical density functions for stock log returns

Before proceeding with option pricing tests, let us examine the distribution of log returns for two stocks, GOOG and DF. For each stock, we assemble a time series Z of daily log returns for the 300 days prior to June 10, 2011. We fit a normal distribution to the time series using the sample mean and variance of Z. We also apply the Expectation Maximization (EM) algorithm to fit a mixture of two normals to Z. Finally, we use kernel density estimation (KDE) to fit a density $f(z)$ to Z.

In Fig. 4, we plot the three densities for GOOG (respectively, DF) in the left (respectively, right) panel. For GOOG, the mixture of two normals fits the KDE density better than the single normal, especially at the peak of the distribution and the region near the peak. For DF, the agreement between the KDE density and the mixture of two normals is even more pronounced. The single normal does not fit nearly as well.

We conclude that, at least for these two stocks, the mixture of two normal distributions fits much better than a single normal. We test this for all stocks in the following way. For each stock, we take the time series of daily log returns and fit (i) a single normal and (ii) a mixture of two normals. After fitting, we calculate the BIC-penalized likelihood for both (i) and (ii). The BIC penalty term accounts for the fact that the mixture has five parameters instead

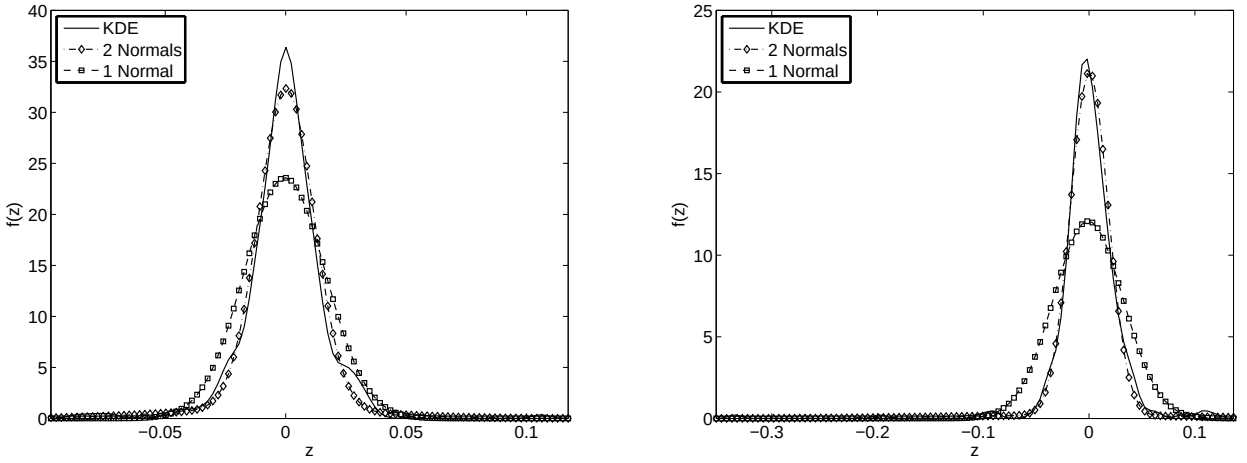

Figure 4: We fit three p.d.f.'s to daily log return time series for GOOG (left) and DF (right). The p.d.f's are a kernel density estimate (KDE), a mixture of two normals, and a single normal. The results show that the mixture of two normals more closely matches the KDE density. See Section 6.2 for more details.

of just two parameters for the single normal.

We find that for 71 out of the 89 total stocks, the BIC-penalized likelihood is larger for the normal mixture distribution. From a model selection point of view, this indicates that the normal mixture is a better choice for modeling log return time series.

6.3 Comparing model and market option prices

We now test the models MT Naive and MT Reg, introduced in Section 6.1, against both Black-Scholes model prices and market prices of options. We collected from Yahoo! Finance 11 days of market prices for options on 89 non-dividend-paying stocks from the S&P 500. In what follows, we refer to the average of the bid and ask prices as the market price of the option.

For MT Reg, the previous day's option prices are required to train the model. Hence with 11 days of options data, we can make a fair comparison between model and market prices for the final 10 days. For these same 10 days, we also compute option prices using the MT Naive and the Black-Scholes models.

Here is how we compute the error on each day. Suppose we have fixed the stock symbol θ and we focus on one particular expiration date τ. Then there will be call options at, say, k different strikes; let $\mathbf{C}^{\text{market}}$, a vector of length k, denote the market prices of these call options. We compute the mean of the absolute values of the relative errors between market and model prices:

$$E^{\text{model}}(\theta, \tau) = \frac{1}{k} \sum_{i=1}^{k} \left| \frac{C_i^{\text{market}} - C_i^{\text{model}}}{C_i^{\text{market}}} \right|,$$

where "model" can take the values B-S (Black-Scholes), MT Reg, or MT Naive. We choose this metric because we are concerned with the percentage errors made in pricing each option that is traded. Other error metrics, such as RMS absolute error in units of dollars, assign lower importance to mispricing options that are worth less.

We then average $E^{\text{model}}(\theta, \tau)$ over all possible expirations τ to obtain the mean error $E^{\text{model}}(\theta)$ committed by the

	σ_*^+				σ_*^-			
	One parameter		Three parameters		One parameter		Three parameters	
	L^2	Tree	L^2	Tree	L^2	Tree	L^2	Tree
10 Jun 2011	0.7582	0.9106	0.7603	0.9167	0.2133	0.4981	0.2249	0.6776
13 Jun 2011	0.7387	0.8744	0.7345	0.8975	0.1529	0.6690	0.1668	0.6485
14 Jun 2011	0.7477	0.8735	0.7478	0.8997	0.1190	0.6417	0.1520	0.5868
15 Jun 2011	0.7279	0.8632	0.7229	0.9047	0.1885	0.6821	0.2051	0.7440
16 Jun 2011	0.7391	0.8842	0.7415	0.9262	0.1339	0.6446	0.1450	0.6551
17 Jun 2011	0.7661	0.8753	0.7778	0.8703	0.3015	0.6857	0.2965	0.7112
20 Jun 2011	0.7696	0.8951	0.7647	0.9017	0.0938	0.6386	0.1547	0.6649
21 Jun 2011	0.7968	0.9260	0.7932	0.9431	0.0286	0.4682	0.0218	0.5350
22 Jun 2011	0.7878	0.9185	0.7858	0.9255	0.0775	0.5495	0.1047	0.5794
23 Jun 2011	0.8323	0.9218	0.8297	0.9331	0.0435	0.4923	0.1034	0.6306
24 Jun 2011	0.8185	0.9022	0.8158	0.9160	0.0724	0.5093	0.0903	0.5303

Table 2: Adjusted R^2 values for the linear models (33) and (34) are given in the L^2 subcolumns with respective column headings "One parameter" and "Three parameters." Adjusted R^2 values for the tree models (35) and (36) are given in the Tree subcolumns with respective column headings "One parameter" and "Three parameters." Note that a reasonable fit for σ_*^- is provided only by Tree models; moreover, the Tree model with three parameters is the best.

	All Symbols			BIC Symbols		
	Black-Scholes	MT Naive	MT Reg	Black-Scholes	MT Naive	MT Reg
13 Jun 2011	0.2026	0.1395	0.1267	0.2217	0.1419	0.1186
14 Jun 2011	0.2158	0.1394	0.1276	0.2378	0.1433	0.1225
15 Jun 2011	0.1839	0.1346	0.1383	0.1988	0.1352	0.1279
16 Jun 2011	0.1732	0.1373	0.1327	0.1854	0.1387	0.1331
17 Jun 2011	0.1637	0.1411	0.1277	0.1741	0.1435	0.1293
20 Jun 2011	0.1947	0.1397	0.1322	0.2110	0.1408	0.1296
21 Jun 2011	0.1977	0.1274	0.1214	0.2182	0.1316	0.1242
22 Jun 2011	0.1923	0.1294	0.1254	0.2129	0.1349	0.1291
23 Jun 2011	0.1830	0.1197	0.1153	0.2017	0.1234	0.1158
24 Jun 2011	0.1685	0.1297	0.1348	0.1824	0.1315	0.1300

Table 3: For each of 10 days of testing, we record the mean of $E^{\text{model}}(\theta)$ for each of three models. For the columns with heading "All Symbols," the mean is taken over all 89 symbols θ, while for the columns with heading "BIC Symbols," the mean is taken over 71 symbols θ for which BIC model selection chooses a normal mixture distribution.

	All Symbols			BIC Symbols		
	Black-Scholes	MT Naive	MT Reg	Black-Scholes	MT Naive	MT Reg
13 Jun 2011	0.0188	0.0047	0.0043	0.0208	0.0050	0.0038
14 Jun 2011	0.0259	0.0064	0.0036	0.0287	0.0070	0.0030
15 Jun 2011	0.0169	0.0049	0.0053	0.0191	0.0052	0.0042
16 Jun 2011	0.0162	0.0058	0.0044	0.0188	0.0064	0.0046
17 Jun 2011	0.0171	0.0082	0.0051	0.0201	0.0094	0.0056
20 Jun 2011	0.0261	0.0079	0.0058	0.0298	0.0085	0.0062
21 Jun 2011	0.0226	0.0060	0.0051	0.0251	0.0065	0.0058
22 Jun 2011	0.0219	0.0058	0.0062	0.0246	0.0064	0.0068
23 Jun 2011	0.0176	0.0036	0.0043	0.0195	0.0039	0.0047
24 Jun 2011	0.0126	0.0041	0.0051	0.0142	0.0044	0.0044

Table 4: For each of 10 days of testing, we record the variance of $E^{\text{model}}(\theta)$ for each of three models. The column headings "All Symbols" and "BIC Symbols" denote the same set of symbols described in Table 3.

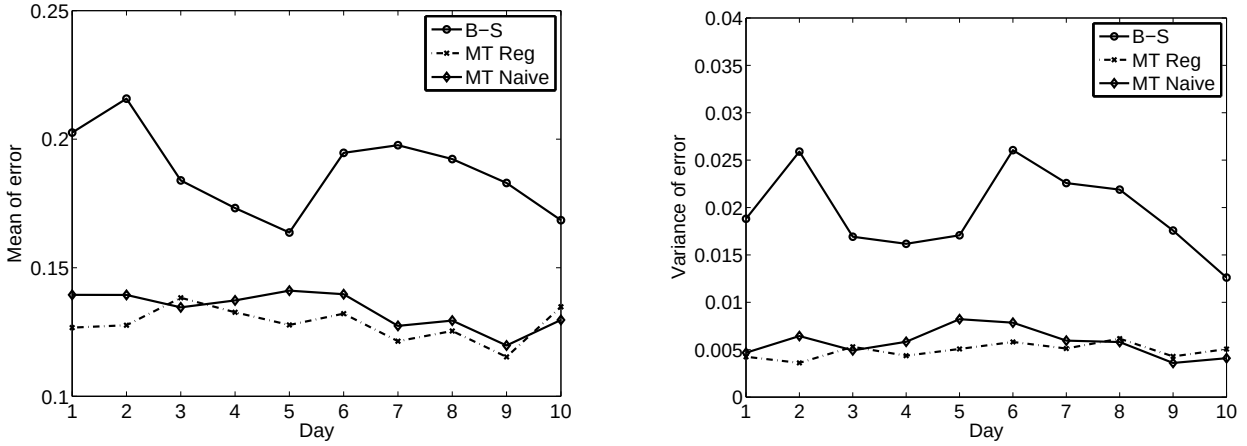

Figure 5: The left (respectively, right) panel shows E^{model} (respectively, $\text{Var}[E^{\text{model}}(\theta)]$) for each of 10 days of testing and each of the three models B-S, MT Reg, and MT Naive. See Section 6.3 for more details.

model for the symbol θ. Finally, we average over all symbols θ to obtain the mean error E^{model} committed by the model. Through all of this, E has the units of fractional error, i.e., $100 \times E$ has units of percentage error.

In the left panel of Fig. 5, we plot E^{model} for each of the 10 days of testing, and for each of the three models. The values that are plotted are also given in Table 3 under the heading "All Symbols." The values that are plotted under the heading "BIC Symbols" are averages of $E^{\text{model}}(\theta)$ over those 71 symbols θ for which BIC selects a normal mixture distribution for the log return time series—see Section 6.2 for more details.

In the right panel of Fig. 5, we plot the variance $\text{Var}[E^{\text{model}}(\theta)]$ for each of the 10 days of testing, and for each of the three models. The values that are plotted are also given in Table 4 under the heading "All Symbols." The values that are plotted under the heading "BIC Symbols" are variances $\text{Var}[E^{\text{model}}(\theta)]$ over those 71 symbols θ for which BIC selects a normal mixture distribution for the log return time series.

Fig. 5 shows that both the mean and the variance of the MT model's errors are less than the B-S model's errors over all 10 days of testing. The small and nearly constant variance of the MT model's errors hints that the method is robust and would fare well over a much longer period of testing. In future work, we intend to pursue exactly such a test.

Tables 3 and 4 also show that, across all days of testing, the MT models perform better than the B-S model. Additionally, we see that using the MT models for symbols for which BIC model selection selects a single normal density does not incur any special penalty. However, if one examines the B-S columns in these tables, one finds that the B-S model does perform noticeably worse on symbols for which BIC model selection chooses a mixture model.

Another visualization of the errors committed by the MT Reg model is provided in Fig. 6. Here we have 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Reg}}(\theta))$. On all of the scatterplots, the vertical axis has been truncated at 0.5, which is sufficient to contain all the points. The horizontal axis has twice the range to account for the errors made by the B-S model. Clearly, the errors made by the MT Reg model are much less dispersed in space than those made by the B-S model. We plot a line of slope one to show that the majority of the 89 points lies below the line, i.e., the MT Reg model's error is less than the B-S model's error for the majority of symbols θ.

The same type of visualization of errors for the MT Naive model is provided in Fig. 7. Again we have 10

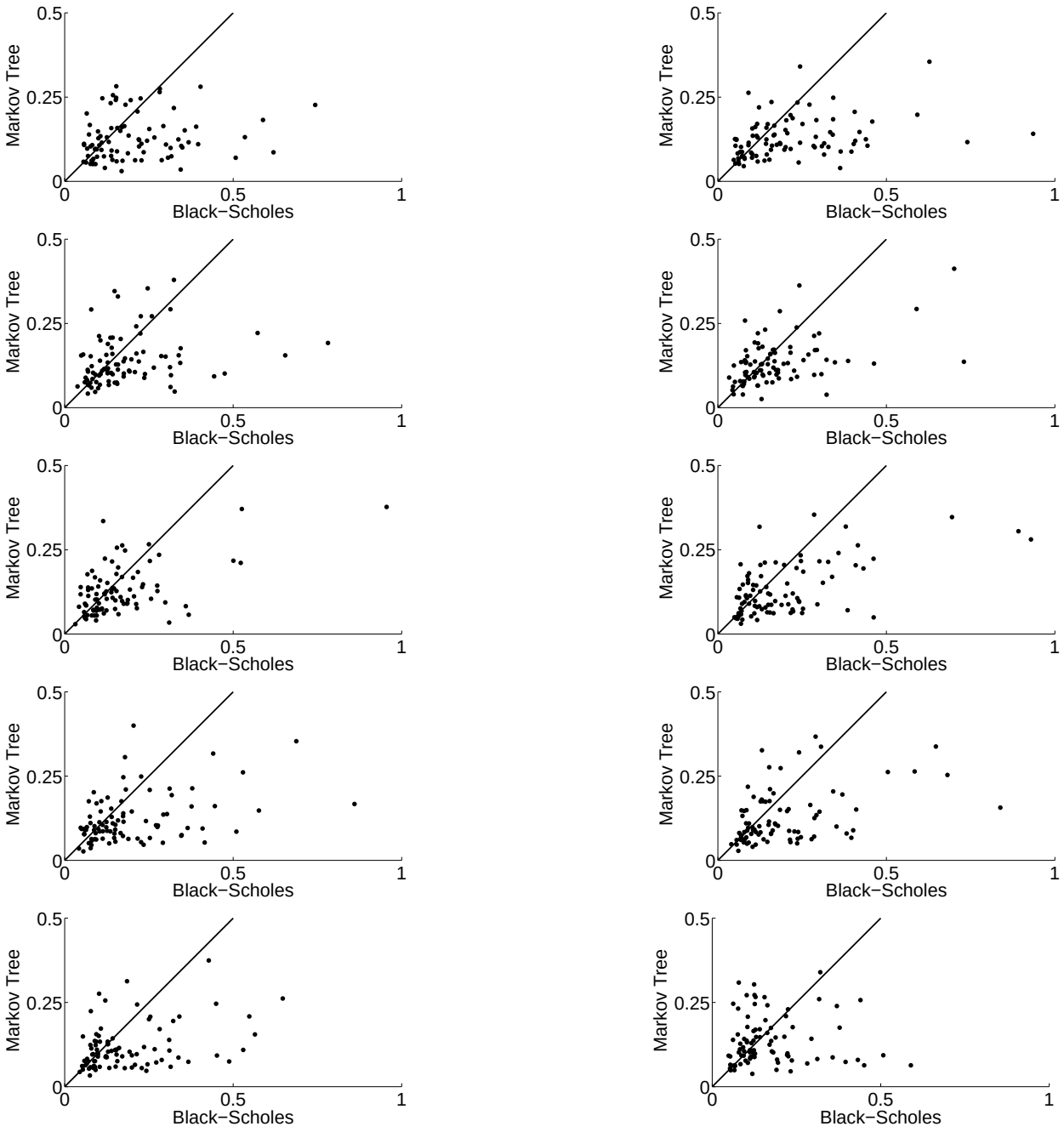

Figure 6: We give 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Reg}}(\theta))$. The majority of the points lie below the line of slope one. The B-S model's errors are larger and more dispersed than the MT model's errors. See Section 6.3 for more details.

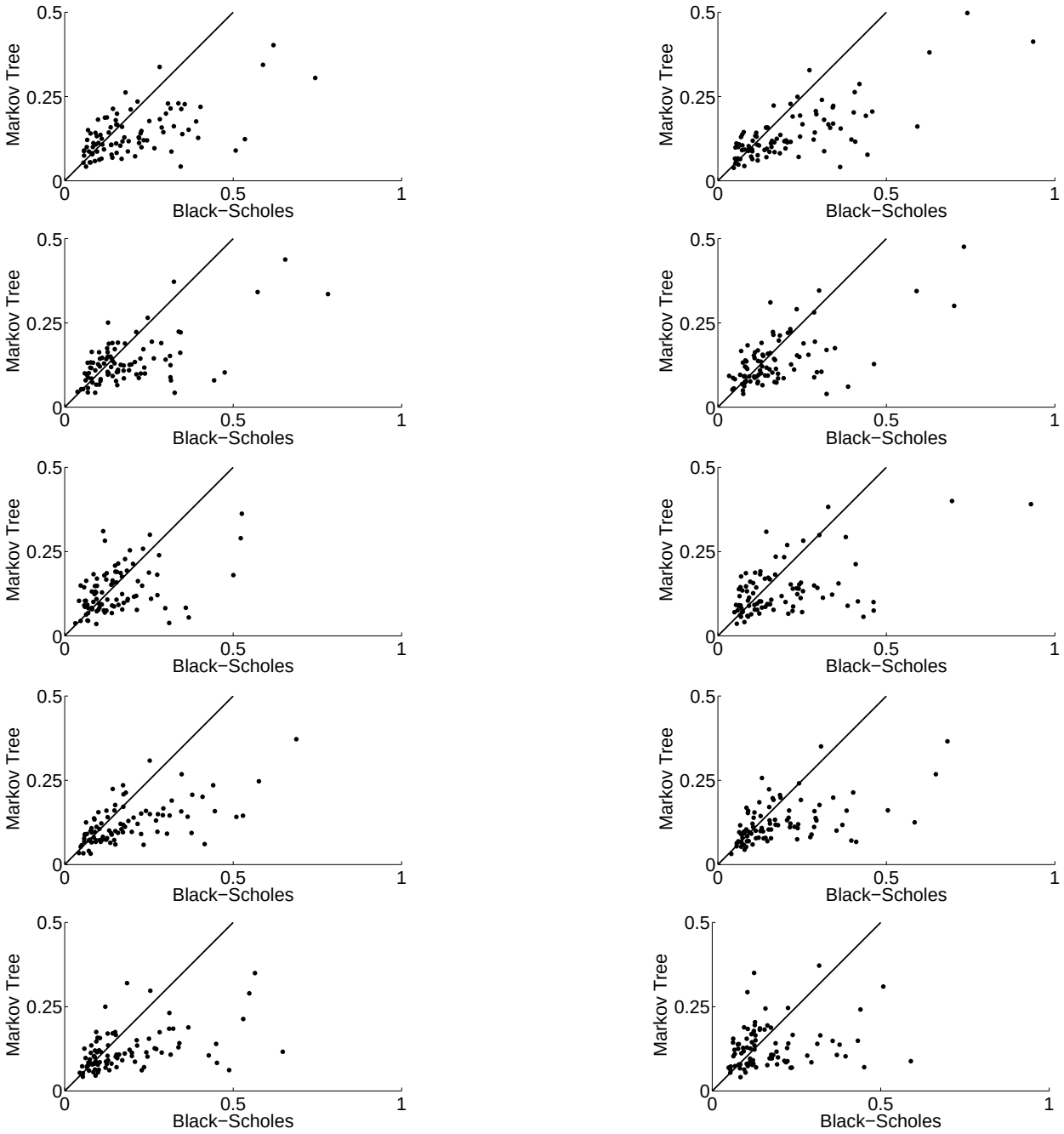

Figure 7: We give 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Naive}}(\theta))$. The majority of the points lie below the line of slope one. The B-S model's errors are larger and more dispersed than the MT model's errors. See Section 6.3 for more details.

scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Naive}}(\theta))$. The performance of the MT Naive model is not quite as sharp as the MT Reg model, but the same general conclusions from the previous paragraph apply.

References

Aingworth, D. D., Das, S. R., and Motwani, R. (2006). A simple approach for pricing equity options with Markov switching state variables. *Quantitative Finance*, 6(2):95–105.

Appleby, J. A. D., Daniels, J. A., and Krol, K. (2012a). A Black–Scholes model with long memory. URL.

Appleby, J. A. D., Riedle, M., and Swords, C. (2012b). Bubbles and crashes in a Black-Scholes model with delay. *Finance and Stochastics*. URL.

Arriojas, M., Hu, Y., Mohammed, S.-E., and Pap, G. (2007). A delayed Black and Scholes formula. *Stochastic Analysis and Applications*, 25(2):471–492.

Barone-Adesi, G. (1985). Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis*, 20(3):299–313.

Basu, A. and Ghosh, M. K. (2009). Asymptotic analysis of option pricing in a Markov modulated market. *Operations Research Letters*, 37(6):415–419.

Behr, A. and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalised logF and generalised hyperbolic models. *Annals of Finance*, 5(1):49–68.

Bhat, H. S. and Kumar, N. (2010). Markov tree options pricing. In *Proceedings of the Fourth SIAM Conference on Mathematics for Industry (MI09)*, pages 162–173, San Francisco, CA. URL.

Bouchaud, J. P. and Potters, M. (2003). *Theory of Financial Risk and Derivative Pricing: from Statistical Physics to Risk Management*. Cambridge University Press.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA.

Brigo, D. and Mercurio, F. (2002). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5(4):427–446.

Buckley, I., Saunders, D., and Seco, L. (2008). Portfolio optimization when asset returns have the Gaussian mixture distribution. *European Journal of Operational Research*, 185(3):1434–1461.

Cai, N. and Kou, S. G. (2011). Option pricing under a mixed-exponential jump diffusion model. *Management Science*, 57(11):2067–2081.

Camara, A. and Li, W. (2008). Jump-Diffusion Option Pricing without IID Jumps. *SSRN eLibrary*. URL.

Campbell, J. Y., Lo, A. W. C., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Chang, M.-H., Pang, T., and Pemy, M. (2010). An approximation scheme for Black-Scholes equations with delays. *Journal of Systems Science and Complexity*, 23(3):438–455.

Chang, M.-H., Pang, T., and Yang, Y. P. (2011). A stochastic portfolio optimization model with bounded memory. *Mathematics of Operations Research*, 36(4):604–619.

Chang, M.-H. and Youree, R. K. (2007). Infinite-dimensional Black-Scholes equation with hereditary structure. *Applied Mathematics and Optimization*, 56(3):395–424.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236.

D'Amico, G., Janssen, J., and Manca, R. (2009). European and American options: The semi-Markov case. *Physica A*, 388(15-16):3181–3194.

Ding, Z., Granger, C. W. J., and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299.

Fama, E. F. (1970). Efficient capital markets—A review of theory and empirical work. *Journal of Finance*, 25(2):383–423.

Fielitz, B. D. (1975). On the stationarity of transition probability matrices of common stocks. *The Journal of Financial and Quantitative Analysis*, 10(2):327–339.

Fielitz, B. D. and Bhargava, T. N. (1973). The behavior of stock-price relatives—a Markovian analysis. *Operations Research*, 21(6):1183–1199.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Heston, S. L. and Nandi, S. (2000). A closed-form GARCH option valuation model. *The Review of Financial Studies*, 13(3):585–625.

Hull, J. C. (2009). *Options, Futures and Other Derivatives*. Prentice Hall finance series. Prentice Hall.

Inverarity, G. W. (2003). Numerically inverting a class of singular Fourier transforms: theory and application to mountain waves. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 459(2033):1153–1170.

Jackwerth, J. C. (1999). Option implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, pages 66–82.

Janssen, J., Manca, R., and Biase, G. D. (1997). Markov and semi-Markov option pricing models with arbitrage possibility. *Applied Stochastic Models and Data Analysis*, 13:103–113.

Kazmerchuk, Y., Swishchuk, A., and Wu, J. (2007). The pricing of options for securities markets with delayed response. *Mathematics and Computers in Simulation*, 75:69–79.

Kendall, M. G. (1944). *The Advanced Theory of Statistics*. Vol. I. J. B. Lippincott Co., Philadelphia.

Kon, S. J. (1984). Models of stock returns—a comparison. *Journal of Finance*, 39(1):147–165.

Lighthill, M. (1958). *Introduction to Fourier Analysis and Generalised Functions*. Cambridge monographs on mechanics and applied mathematics. Cambridge University Press.

Longin, F. (2005). The choice of the distribution of asset returns: How extreme value theory can help? *Journal of Banking & Finance*, 29(4):1017–1035.

Mamon, R. S. and Rodrigo, M. R. (2005). Explicit solutions to European options in a regime-switching economy. *Operations Research Letters*, 33(6):581–586.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4):308–313.

Niederhoffer, V. and Osborne, M. F. M. (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, 61(316):897–916.

Ramponi, A. (2011). Mixture dynamics and regime switching diffusions with application to option pricing. *Methodology and Computing in Applied Probability*, 13(2):349–368.

Ritchey, R. J. (1990). Call option valuation for discrete normal mixtures. *Journal of Financial Research*, 13(4):285–296.

Rudnick, J. A. and Gaspari, G. D. (2004). *Elements of the Random Walk: An Introduction for Advanced Students and Researchers*. Cambridge University Press.

Sewell, M. (2011). Characterization of financial time series. Technical Report RN/11/01, University College London, London. URL.

Shreve, S. E. (2004). *Stochastic Calculus for Finance. I.* Springer Finance. Springer-Verlag, New York.

Swords, C. and Appleby, J. A. D. (2010). *Stochatic Delay Difference and Differential Equations*. Lambert Academic Publishing.

Tan, K. and Chu, M. (2012). Estimation of portfolio return and Value at Risk using a class of Gaussian mixture distributions. *The International Journal of Business and Finance Research*, 6(1):97–107.

Taylor, S. J. (2007). *Introduction to Asset Price Dynamics, Volatility, and Prediction*. Princeton University Press.

Wu M., Huang N., and Zhao C. (2008). European option pricing with time delay. In *2008 Chinese Control Conference (CCC)*, pages 589–93, Piscataway, NJ, USA. IEEE.

A Expressions for Normal Mixture Distribution Parameters

We list expressions for the constants μ_1, σ_1, μ_2, and σ_2 that appear in the normal mixture distribution (27).

$$
\begin{aligned}
\mu_1 &= \tilde{S}_0 - l_u + \frac{l_1(q^+ - 1)(q^- + q^+) + l_2(q^-(q^- - 3q^+ + 3) + q^+ - 1)}{(q^- - q^+ + 1)^2} \\
&\quad + \frac{(n-1)(l_1 q^-(1 - 2q^+) + l_2(2q^- - 1)(q^+ - 1))}{q^- - q^+ + 1} \\
\sigma_1^2 &= -(n-1)q^-(q^+ - 1)\Big[q^+\left(l_1^2(4q^-(q^- + 2) - 3) - 2l_1 l_2(4q^-(q^- + 2) - 9) + l_2^2(4q^-(q^- + 2) - 11)\right) \\
&\quad - 4(q^- - 1){q^+}^2(l_1 - l_2)^2 - (q^- - 1)(l_1 - 3l_2)^2\Big](q^- - q^+ + 1)^{-3} \\
&\quad - (q^+ - 1)\Big[l_1^2\left({q^-}^3 + {q^-}^2(2q^+ - 1) + q^-(q^+(9q^+ - 8) + 1) + (q^+ - 1)q^+\right) \\
&\quad - 2l_1 l_2\left({q^-}^3 + {q^-}^2(6q^+ - 5) + q^-(5(q^+ - 2)q^+ + 3) - {q^+}^2 + q^+\right) \\
&\quad + l_2^2\left({q^-}^3 + {q^-}^2(10q^+ - 13) + q^-((q^+ - 12)q^+ + 13) + (q^+ - 1)q^+\right)\Big](q^- - q^+ + 1)^{-4} \\
\mu_2 &= \tilde{S}_0 + l_u + \frac{l_1\left(q^-(3q^+ - 2) - (q^+ - 1)^2\right) - l_2 q^-(q^- + q^+ - 2)}{(q^- - q^+ + 1)^2} \\
&\quad + \frac{(n-1)(l_1 q^-(1 - 2q^+) + l_2(2q^- - 1)(q^+ - 1))}{q^- - q^+ + 1} \\
\sigma_2^2 &= -(n-1)q^-(q^+ - 1)\Big[q^+\left(l_1^2(4q^-(q^- + 2) - 3) - 2l_1 l_2(4q^-(q^- + 2) - 9) + l_2^2(4q^-(q^- + 2) - 11)\right) \\
&\quad - 4(q^- - 1){q^+}^2(l_1 - l_2)^2 - (q^- - 1)(l_1 - 3l_2)^2\Big](q^- - q^+ + 1)^{-3} \\
&\quad - q^-\Big[l_1^2\left({q^-}^2(q^+ - 2) + q^-(10(q^+ - 1)q^+ + 1) + {q^+}^3 - q^+\right) \\
&\quad - 2l_1 l_2\left({q^-}^2(5q^+ - 4) + q^-(6(q^+ - 2)q^+ + 5) + (q^+ - 3)(q^+ - 1)q^+\right) \\
&\quad + l_2^2\left({q^-}^2(9q^+ - 10) + q^-(2(q^+ - 7)q^+ + 13) + (q^+ - 1)((q^+ - 3)q^+ + 4)\right)\Big](q^- - q^+ + 1)^{-4}
\end{aligned}
$$