UNIVERSITY OF CALIFORNIA, MERCED

PH.D. DISSERTATION

Modeling Dependence in Data: Options Pricing and Random Walks

Nitesh Kumar

A dissertation submitted
in partial fulfillment of the requirements for the degree

Doctor of Philosophy in Applied Mathematics

March, 2013

UNIVERSITY OF CALIFORNIA, MERCED

Graduate Division

This is to certify that I have examined a copy of a dissertation by

Nitesh Kumar

and found it satisfactory in all respects, and that any and all revisions required by the examining committee have been made.

Faculty Advisor:

Harish S. Bhat

Committee Members:

Arnold D. Kim

Roummel F. Marcia

Applied Mathematics Graduate Studies Chair:

Boaz Ilan

Arnold D. Kim

Date

Contents

Acknowledgements

This thesis and my sanity would not be intact if it had not been for my advisor Harish Bhat. I thank Harish for taking me under his wings early in Fall, 2008 when I was still trying to find my feet in academia and research. His uncanny ability to recognize both my failures and successes well before I could is what has brought this thesis to its timely conclusion. I thank Harish for the past five years of advising and friendship and many more years to come.

My next main academic sources of influence have been my committee members Arnold Kim and Roummel Marcia who have provided me with constant encouragement all through my graduate career. Their academic jokes and anecdotes will surely keep me smiling for long.

I thank Francois Blanchette and Lei Yue for making me and other international students feel at home through yearly Christmas celebrations at their place. I also thank Francois for numerous times he listened to me and helped me out with my indecisions.

At UC Merced, I learnt much of the mathematics that allowed me to undertake this thesis and complete it. To this I am forever thankful to the professors who taught me here– Boaz Ilan, Mayya Tokman, and Arnold Kim. I also thank Carrie King for her valuable support over the years and making academic paper work less tedious.

Through my years in graduate school, I also made long lasting friends who made my life at UC Merced fun filled and exciting. I thank Jane for inspiring me when disappointment started to raise its ugly head. I also thank her for sweating it out together seeing the funny side of academic proceedings. I also thank Derya for years of companionship and fruitful discussions about research in particular and life in general. She ensured I stayed logical and reasonable at all times despite my attempts at the exact opposite. I have learnt more from these two ladies than I could have ever hoped to learn from any other colleagues or friends.

I also take this opportunity to thank Garnet, my friend, academic brother, and roommate of three years. His cooking ensured that I never missed home cooked Indian food. I also thank him for insightful discussions over my research, and his quick, often unforgiving feedback.

Finally, I thank my parents who have inspired me to chase my dreams. I also thank my brother for kindling in me the passion for mathematics and research at an early age. He has done more than he will ever take credit for.

Modeling Dependence in Data: Options Pricing and Random Walks

by

Nitesh Kumar

University of California, Merced, 2013

Prof. Harish S. Bhat, Chair

ABSTRACT OF THE DISSERTATION

In this thesis, we propose the Markov tree option pricing model and subject it to large-scale empirical tests against market options and equity data to quantify its pricing and hedging performances.

We begin by proposing a tree model that explicitly accounts for the dependence observed in the log-returns of underlying asset prices. The dynamics of the Markov tree model is explained together with implementation notes that enable exact calculation of the probability mass function of the Markov tree process. We also show that the tree model operates in the framework of arbitrage free option pricing.

Next, we show how the discrete Markov tree process can be viewed as a generalized persistent random walk and demonstrate how to approximate it by a mixture of two normals. This derivation enables us to obtain a closed form pricing formula for the European call option allowing for faster calibration using market option data. We then empirically test both the pricing as well as the hedging performance of the Markov tree model against the Black-Scholes and the Heston's stochastic volatility models establishing its superior hedging performance. Additionally, we also analyze different regression based techniques to estimate the parameters in the Markov tree model that obtain increasingly better hedging results. We also lay down statistical procedures to rigorously analyze the hedging performance of any option pricing model.

We then generalize the Markov tree process and explore its relation with the generalized delayed random walk. In doing so, we develop a spectral method for computing the probability density function for delayed random walks; for such problems, the spectral method we propose is exact to machine precision and faster than existing methods. In conjunction with step function approximation and the weak Euler-Maruyama discretization, the spectral method can be applied to nonlinear stochastic delay differential equations. We carry out tests for a particular nonlinear SDDE that shows that this method captures the solution without the need for Monte Carlo sampling.

Chapter 1

Introduction

Substantial development in computational power and storage space have enabled researchers to analyze data more than ever before. Financial data, specifically, is not just readily available but requires little to no cleaning unlike the data sets appearing in other fields. This recent development has made it possible for researchers to test the assumptions made in financial models and also empirically evaluate the performance of such models. In this dissertation, we use data driven mathematical models to solve the option pricing problem that arises in mathematical finance and further generalize these techniques to develop methods to solve problems arising in neuroscience and biology.

1.1 Brief Review of the Option Pricing Problem and Models

The fundamental problem in mathematical finance is the option pricing problem that, at the face of it involves calculating the fair price of an option given the option parameters and the spot (current) price of its underlying. A European style call (put) option is a financial contract that gives the buyer the *option* (i.e. the right without the obligation) to buy (sell) the underlying instrument at a given price called the *strike* price at a particular date in future called *expiry*. If the buyer of the European call (put) option chooses to exercise the right, then the seller is under the obligation to fulfill it by selling (buying) the underlying instrument at the strike price when the option expires.

A closed form option price is highly desirable– apart from calculating the price of the option, such a formula enables traders and practitioners to create a risk free portfolio and hedge market risk. The 1987 crash and the current financial crisis (2008–present) have brought to light poor risk management that, in turn, are consequences of poor option pricing models. Since mathematical models for option pricing form the basis of risk management, the current crisis has revealed more than ever before that the option pricing problem is far from solved.

The basic ingredient in any option pricing model is the stochastic process for the asset price.

Hence, it is not surprising that much of the research in option pricing is driven by the quest to incorporate a stochastic process that agrees well with the observed asset prices. The first popular mathematical model to price options, the Black-Scholes (BS) model used the now famous no arbitrage argument to arrive at the fair price of a European option Black and Scholes (1973). In their work, they modeled the stock price process as a geometric Brownian motion (gBm) to obtain a closed form expression for the European option price. The proposition that the stock price process is actually a gBm has met with a lot of criticism mainly due to its inability to capture tail behaviour in the stock returns observed during a crash (MacKenzie, 2004). Empirical research has also validated that the observed log returns of stocks have 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). While these inconsistencies remain, the BS model has several upshots. First, the BS model gives a simple closed form analytical expression for the option price enabling faster computation of the option price. Second, the Black-Scholes model is a one parameter model keeping its calibration to market option prices tractable and computationally inexpensive. Third, the BS model is based on the principles of the well accepted arbitrage free pricing theory. Finally, the binomial option pricing model that converges to the BS model is a tree model that facilitates understanding of the stock price movements at discrete time steps[1] in the BS model Cox et al. (1979). Option pricing models that attempt to incorporate the observed features noted above in the log returns of the stock price process are increasingly complicated and often fail to retain the upshots of the BS model.

Another feature of the gBm assumption in the BS model is that the log returns of the stock price process are assumed to be independently and identically distributed (IID). This can be easily understood by studying the binomial tree model that converges to the BS model in the limit as the time duration of each time step in the tree goes to zero. In the binomial tree model, the log return of the stock price process is assumed to follow a simple biased random walk where both the increments and the probabilities associated with the increments are constant. Hence, it is easy to note that every step of the binomial tree is independent of its previous steps. While the deviation of the observed log returns from the normal distribution has been well studied in literature, the IID assumption has been rarely addressed before. Empirical studies on markets on the other hand indicate that the daily log returns of stocks are *not* generated from an IID process (Ding et al., 1993; Lo and MacKinlay, 1988)[2] The IID hypothesis is extensively tested by studying the auto-correlation of the transformed time series $\{|X_n|^\lambda\}, \lambda \in 1, 2, \ldots$[3] (Ding et al., 1993). Significant

[1] In reality the stock price process only changes at discrete time intervals.

[2] Although the log returns time series is not IID, there is little predictability due to zero autocorrelation of the time series observed (French and Roll, 1986; Lo and MacKinlay, 1990; Blair et al., 2002).

[3] If $\{S_t\}_{t=0}^n$ is the stock price time series at equispaced intervals of time, then the log returns time series $\{X_t\}$ is defined as $X_t = \log(S_{t+1}/S_t)$.

positive autocorrelation at lags 1-5 of the process $\{|X_n|^\lambda\}$ is enough to reject the IID hypothesis for the original log returns time series $\{X_n\}$ (Taylor, 2007, Chap. 4). Building upon the binomial model, multinomial models like a trinomial model and its extension (Kamrad and Ritchken, 1991), a pentanomial model that accounts for skewness and kurtosis observed in the underlying asset (Primbs et al., 2007), and an octanomial tree for the stock price (Leisen, 2000) have been proposed. The quadrinomial tree model (Florescu and Viens, 2008) accounts for stochastic volatility in its tree model, like its predecessor (Aingworth et al., 2006). Even though these models provide a simple understanding of the stochastic process assumed for the stock price by accounting for different features observed in the stock price process, none of these models explicitly account for the non-IID behaviour reflected in the observed log returns. In Chapter 2, we propose the Markov tree model for option pricing that explicitly accounts for the non-IID behaviour observed in the log returns through a simple tree model in the framework of classical arbitrage free pricing. In the same Chapter, we also propose a method to test dependence in observed log returns of the stock price process. Testing the Markov tree model against market options data on six different stocks for a period of 45 trading days from the Paris stock exchange leads us to conclude that explicitly accounting for the non IID behaviour leads to better pricing performance when compared to the Black-Scholes model.

While practitioners find it easy to understand the discrete time evolution of the stock price process through a tree model, they are numerically expensive to calibrate against market data by virtue of being a discrete model. Hence, a continuous model that leads to a closed form European option price is highly desirable. The Black-Scholes model discussed above is one such model that provides a closed from European call option price. It 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 (i) normality of daily log returns, and (ii) independence of increments. In Chapter 3, we examine in detail, both theoretically and empirically, the MT model in which both assumptions are removed. By construction, the MT model accounts for the serial dependence of log returns. As we show in Chapter 3, the distribution generated by the MT model is closely approximated by a mixture of normals leading to a closed form pricing formula for the European call option. Through 10 days of empirical testing on options traded on 89 of the S&P 100 components, we establish that the MT model prices are closer on average to the market prices than the BS model prices.

The Black-Scholes option price is extremely close to the true market option price for at the money (ATM) options. However, for in the money (ITM) options and out of the money (OTM) options, the BS price does not agree well with the market option price. The BS formula can be easily inverted to calculate the volatility (commonly referred to as the implied volatility) for which

BS formula gives the exact market price. The ITM and OTM options are known to have higher implied volatility than the ATM options, a phenomenon commonly known as the volatility smile. This implied volatility is also known to change with the expiration and is referred to as the term-structure. These observations regarding the smile and the term structure have cast significant doubt over the constant volatility assumption in the Black-Scholes model. As a result, a number of option pricing models have emerged that treat the volatility as a stochastic random variable. Typically, the volatility is modeled as a stochastic process that is governed by a Brownian motion that is correlated with the Brownian motion driving the stochastic process for the stock price (Heston, 1993; Hagan et al., 2002). These stochastic processes fit the market option prices very well but do not confirm to the classical arbitrage free pricing framework. Generalizations of these stochastic volatility model involve, incorporating jumps in the stock price process, assuming a stochastic process for the interest rates, and, combining both these features Bakshi et al. (1997). Empirical tests Bakshi et al. (1997); Kaeck (2012) against market data shows that the stochastic volatility models produce better pricing than the Black-Scholes model. They also conclude that generalizations of the stochastic volatility models noted above do not necessarily improve the hedging performance of Heston's stochastic volatility model. While the option price can always be looked up from the market a closed form option pricing formula is required to create a risk-free portfolio and hedge market risk. It is for this reason that any option pricing model should be evaluated through its hedging performance and its ability to hedge market risk Kaeck (2012).

In Chapter 4, we subject the Markov tree model to empirical tests against the most popular stochastic volatility model, the Heston's stochastic volatility model Heston (1993). Using over three years of Paris LIFFE equity options data, we conclude that the Markov tree model outperforms both the Black-Scholes model and Heston's stochastic volatility model in terms of out-of-sample hedging performance. Empirical tests on two years of S&P 500 index options confirm the conclusions made from the Paris LIFFE equity options.

The Markov tree stochastic process captures key features of the stock price process that in turn lead to better hedging performance than other option pricing models considered in this thesis. In Chapter 5, we seek to generalize the Markov tree stochastic process to model other stochastic processes arising in biology and neuroscience. We show that the generalization of the Markov tree process can be viewed as a delayed random walk and further develop a spectral method to obtain the probability mass function of such a random walk. Through a step function approximation, we show how this numerical method can be used to solve for a stochastic delayed differential equation.

Chapter 2

Markov Tree: Discrete Model

2.1 Introduction

In the Black-Scholes model for the price of a European option, one of the main assumptions is that the price of the underlying asset follows a geometric Brownian motion (Hull, 2009). If S_t is the underlying asset price at time t, one assumes $dS_t = \mu S_t dt + \sigma S_t dW_t$, where μ and σ are constants and W_t is a Brownian motion. For fixed $t > 0$, define $X_n = \log(S_{(n+1)t}/S_{nt})$. Then

$$X_n = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma \left(W_{(n+1)t} - W_{nt}\right).$$

Since W_t is a Brownian motion, $W_{(n+1)t} - W_{nt}$ is normally distributed with mean 0 and variance t. This implies that X_n is normally distributed with mean $(\mu - \sigma^2/2)t$ and variance $\sigma^2 t$, i.e., the distribution of X_n does not depend on n, so each X_n is identically distributed. Moreover, X_{n+1} is independent of X_n, so

$$P(X_n|X_{n-1}) = P(X_n) \tag{2.1}$$

for all positive integers n, and hence the $\{X_n\}$ sequence is IID (independent and identically distributed).

In fact, (2.1) follows from assumptions that are much more general than the geometric Brownian motion assumption; for example, if we take $S_t = S_0 \exp(L_t)$ where L_t is *any* Lévy process, (2.1) still holds. The upshot is that most options pricing models in use—including Black-Scholes, binomial, and most jump-diffusion models—implicitly assume that the daily log returns for any stock are IID. With this in mind, the plan for this Chapter is as follows:

1. We first check whether (2.1) is consistent with real data. To do this, we apply order estimators to log return time series data for European stocks in the CAC-40 index. For several stocks, we find that (2.1) can be rejected in favor of a first-order Markov model for the stock price

Strike	Market	Black-Scholes	Markov Tree
40	34.49	36.57	35.85
48	27.48	29.85	28.09
56	20.90	23.96	20.83
60	17.78	21.36	17.53
64	15.03	18.99	14.53
72	10.00	14.90	9.55
80	6.26	11.60	5.94
88	3.70	8.99	3.53
120	0.32	3.17	0.32
160	0.01	0.87	0.01

Table 2.1: Market and model prices (in €) for a particular European call option on August 24, 2009.

process.

2. We modify the standard binomial tree model to formulate a method for pricing options that is valid when (2.1) is not. We introduce first-order Markov behavior of the underlying asset into the tree, by allowing the jumps of the tree to depend on whether the previous jump was an upward or downward jump. We refer to this model as the Markov Tree (MT) model.

3. Finally, we test the MT model against the standard Black-Scholes model. We find that the MT model's option prices are much closer to market prices than the Black-Scholes model's prices. As a preview of our results, we present Table 2.1, which compares model and market prices on August 24, 2009, for a particular European call option.

2.2 Motivation

Let us discuss Table 2.1 in greater detail. On August 24, 2009, we obtained from euronext.com the end-of-day market prices for European call options for Air Liquide (symbol: AI) expiring in September 2010. We have tabulated the market prices together with prices calculated using the Black-Scholes (Black and Scholes, 1973) model and the MT model introduced in this Chapter. To calculate prices using the Black-Scholes model, we require two parameters, the risk-free interest rate r and the volatility σ. Using standard estimation procedures from empirical data[1], we obtain

[1]We estimate the risk-free rate using the no-arbitrage futures pricing formula $F = Se^{rt}$; here F is the futures price, S is the spot price, and t is the time until expiration of the futures contract. On August 24, 2009, we found that $S = 75.43$ and $F = 75.658$ for the AI future expiring in December 2009, which also gives $t = 84$ trading days $= 0.33$ years. This yields an annualized risk-free rate of $r = 0.0090543$. To estimate the volatility, we start with 252 trading days (or one year) of the adjusted closing price for AI, which we represent as $\{S_1, S_2, \ldots, S_{252}\}$. We then calculate $\hat{\sigma}$, the standard deviation of the log return sequence $\{\log S_2/S_1, \log S_3/S_2, \ldots, \log S_{252}/S_{251}\}$; this yields the annualized volatility $\sigma = \hat{\sigma}\sqrt{252} = 0.41632$. This follows (Hull, 2009, Chap. 13).

$r = 0.00905453$ and $\sigma = 0.41632$. The MT model uses these two parameters together with σ_+ and σ_-, which are the volatilities on days where the stock's return increased (for σ_+) or decreased (for σ_-) relative to the previous day's return[2].

Examining Table 2.1, we find that for a strike of €40, the Black-Scholes model's price is only 6% greater than the market price, but as the strike increases and exceeds the spot price of €75.43, the Black-Scholes model's price diverges considerably. For example, at a strike of €88, the Black-Scholes model's price is 143% greater than the market price. This well-known divergence is usually explained through the dependence of volatility on the strike price. For each strike, one computes the value of the volatility such that the Black-Scholes model price matches the market price. When the resulting implied volatilities are plotted versus strike price, one obtains the classic volatility smile (Hull, 2009, Chap. 16).

We do not dispute that volatility should vary in some way as a function of option strike and time until expiry. However, in the absence of an exact form of this quantitative dependence, we ask: do we know for sure that the discrepancy between Black-Scholes and market prices is due *entirely* to the volatility smile? Our view is that, for certain options, the discrepancy is at least partially due to the market's knowledge that today's returns alter or influence the probability distribution of tomorrow's returns. Unlike commonly used option pricing models, the MT model accounts for this, and as shown in Table 2.1, it is significantly more accurate than Black-Scholes for out-of-the-money options, with no strike-dependent volatilities used for either model. Though the MT model does not provide an analytical formula for the option price, it is computationally tractable thanks to a large amount of recombination in the price tree for the underlying asset. We revisit these implementation issues later in the Chapter.

2.3 Past Work

Before continuing with the plan of the Chapter given in Section 2.1, we discuss relevant past work. A k-th order Markov chain is defined as a sequence $\{Y_n\}_{n\geq 1}$ of random variables such that

$$P(Y_n|Y_{n-1},\dots,Y_1) = P(Y_n|Y_{n-1},\dots,Y_{n-k}).$$

In a k-th order Markov chain, the current state Y_n is allowed to depend only on the past k states. The *order estimation problem* is to take N observations $y_1,\dots,y_N$ generated by a Markov chain (of unknown order) and return an estimate $\hat{k}$ of the chain's order. The estimator is *consistent* if, as the number of observations N goes to infinity, $\hat{k}$ converges to the true order k of the Markov chain.

[2]The parameters $\sigma_\pm$ are calculated in precisely the same way as σ, except that for σ_+ we take the standard deviation of log returns on days when the stock's return increased, while for σ_- we take the standard deviation of log returns on days when the stock's return decreased. This is discussed in greater detail in Section 4.3.

In our work, we make use of the provably consistent BIC order estimator (Csiszár, 2002).

In the context of jump-diffusion models, the IID assumption has been examined recently by Câmara and Li (Camara and Li, 2008), who discuss several empirical studies that have rejected that stock jumps are IID. The focus of their paper is the development of a jump-diffusion options pricing model that does not assume the jumps are IID. Their work differs from ours in two ways: (1) non-IID behavior is modeled only through the jumps (and not through the diffusion) of the jump-diffusion process that the underlying asset is assumed to follow, and (2) the means and variances of the jumps are allowed to be time-varying. By comparison, because the MT model is discrete in time, every stock price path is a sequence of jumps—non-IID behavior is not confined to one part of the model. We make no claims about the limit of the MT model as the number of steps becomes infinite. However, we do assume that the magnitudes of possible jumps remain constant throughout the price tree.

Markov and semi-Markov processes, including processes with finite state spaces, have been used to price options (Janssen et al., 1997; D'Amico et al., 2009). Though these works assume that the log return process $\log(S_t/S_{t-1})$ follows some type of discrete-time Markov or semi-Markov process, the tree models that are proposed differ from the MT model in one important regard: starting from any vertex of the tree, the *magnitudes* of the up and down jumps are always the same. The same is true in models where a Markov chain is used to approximate the true underlying process—see (Duan and Simonato, 2001), for instance. In the MT model, if we start from a vertex such that the jump leading to that vertex *was an upward jump*, then we have different up/down magnitudes as compared with a vertex such that the jump leading to that vertex *was a downward jump*. In other words, the magnitudes of the jumps in the MT model's tree possess the first-order Markov property. This same property distinguishes the MT model from other tree models that involve trinomial, pentanomial, or more general branching at tree vertices—see (Primbs et al., 2007; Yamada and Primbs, 2002).

2.4 Order Estimation: Methodology

Here we describe the methods used to test (2.1) against real data. We begin with a time series $\{s_0, \ldots, s_N\}$ consisting of the adjusted daily closing price of a given stock. We define $x_n = \log(s_n/s_{n-1})$ and obtain the log return time series $\{x_1, \ldots, x_N\}$. Note that each element of this time series is real-valued. To apply Markov order estimation, we must first convert the log return time series into a sequence of symbols drawn from a finite set. As in prior work (McQueen and Thorley, 1991; Tan and Yılmaz, 2002), the simplest way to do this is with just two symbols. We therefore define

$$z_n = \begin{cases} u & x_n \geq 0 \\ d & x_n < 0, \end{cases} \tag{2.2}$$

where the symbols u and d stand for "up" and "down," respectively. Note that this transformation erases the magnitudes of the upward/downward movements of the stock. Now that we have a sequence $\{z_j\}_{j=1}^N$ of u's and d's, we can begin to extract maximum likelihood estimates (MLE's) of Markov chain transition probabilities. Let us describe how this is done.

By the definition given in Section 2.3, a zeroth-order Markov chain is simply a sequence of IID random variables. Since each z_j in our sequence can be in one of only two possible states, if the sequence was generated by a zeroth-order Markov chain, then each z_j was generated by a Bernoulli random variable with only one parameter: $p = P(u)$, the probability of obtaining u. In this case, $1 - p = P(d)$. In this case, we define the zeroth-order log likelihood

$$L_0(p) = n_u \log p + n_d \log(1 - p),$$

where n_u is the number of u's observed in the sequence, $n_d = N - n_u$ is the number of d's observed in the sequence, and N is the total length of the sequence. Solving $dL_0/dp = 0$ for p gives the MLE $\hat{p} = n_u/N$, which is in fact the maximizer of $L_0(p)$. Using this MLE, we can compute the maximum value $L_0(\hat{p})$.

Let us now redo this calculation assuming that the sequence $\{z_j\}_{j=1}^N$ was generated by a first-order Markov chain. Now we require three parameters, $p_1 = P(u)$, $p_2 = P(u|u)$ and $p_3 = P(d|d)$. Note that p_1 is necessary to handle z_1, the first element of the sequence. Also note that $P(d|u) = 1 - p_2$ and $P(u|d) = 1 - p_3$. Putting it all together, we obtain the first-order log likelihood

$$\begin{aligned} L_1(p_1, p_2, p_3) = {} & m \log p_1 + (1 - m) \log(1 - p_1) + n_{uu} \log p_2 + n_{ud} \log(1 - p_2) \\ & + n_{dd} \log p_3 + n_{du} \log(1 - p_3). \end{aligned} \tag{2.3}$$

Here $m = 1$ if $z_1 = u$ and $m = 0$ if $z_1 = d$. The notation $n_{\sigma\tau}$ denotes, for any choice $\sigma, \tau \in \{u, d\}$, the number of times the string $\sigma\tau$ was observed in the sequence. Now solving $\partial L_1/\partial p_j = 0$ for p_j, $j = 1, 2, 3$ yields the MLE's

$$\begin{aligned} \hat{p}_1 &= m \\ \hat{p}_2 &= \frac{n_{uu}}{n_{uu} + n_{ud}} = \frac{n_{uu}}{\text{\# of } u \text{ in first } N - 1 \text{ slots}} \\ \hat{p}_3 &= \frac{n_{dd}}{n_{dd} + n_{du}} = \frac{n_{dd}}{\text{\# of } d \text{ in first } N - 1 \text{ slots}} \end{aligned}$$

		$\hat{k}$		
		0	1	2
k	0	966	33	1
	1	180	818	2
	2	28	116	856

Table 2.2: The $(k, \hat{k})$ entry equals the number of times the BIC order estimator returned an order of $\hat{k}$, when applied to a random sequence of length $N = 500$ generated by a randomly generated Markov chain of true order k. For results in this table, transition probabilities were drawn uniformly from $(0, 1) \subset \mathbb{R}$.

Using these MLE's, we can calculate the maximum value $L_1(\hat{p}_1, \hat{p}_2, \hat{p}_3)$.

Following the same methodology, we can assume that the sequence $\{z_j\}_{j=1}^N$ was generated by a k-th order Markov chain and then write down the k-th order log likelihood function L_k of the unknown Markov transition probabilities $\mathbf{p} = (p_1, \ldots, p_M)$. We plug into L_k the frequencies of different strings found in the actual $\{z_j\}_{j=1}^N$ sequence and then maximize L_k over $\mathbf{p}$. We thereby find both the MLE's $\hat{\mathbf{p}} = (\hat{p}_1, \ldots \hat{p}_M)$ as well as the maximum value of the log likelihood $L_k(\hat{\mathbf{p}})$. The calculations are carried out in the Appendix A, for a sequence that is assumed to be an i-th order Markov chain over Q states. (In the above discussion, we treated only the $Q = 2$ case.)

Armed with this information, we employ the BIC (Bayesian Information Criterion) order estimation method. We calculate $f(j, N) = L_j(\hat{\mathbf{p}}) - 2^{j-1}\log(N)$, looping over values of j from 0 to K. The BIC order estimate $\hat{k}$ equals the value of j that maximizes $f(j, N)$. In the limit where the number of data points is infinite, $N \to \infty$, it has been proven that the BIC estimate converges to the true order of the Markov chain generating the data (Csiszár, 2002).

Though the theoretical results on BIC order estimation with an infinite amount of data are encouraging, they are obviously not strictly applicable to our situation, where the length of the time series is finite. To remedy this, we study the performance of BIC order estimation on finite, synthetic data sets.

For $k \in \{0, 1, 2\}$, we randomly generate transition probabilities for a k-th order Markov chain. Each probability is drawn uniformly from the interval $(0, 1)$. Using this Markov chain, we generate a sequence of length $N = 500$. We apply BIC order estimation to this sequence and thereby obtain an estimate $\hat{k}$ of the Markov chain's order. The results are summarized in Table 2.2. When we apply BIC order estimation, we loop over possible orders $j = 0, 1, 2, \ldots, 8$. However, in no instance do we find that the estimate $\hat{k}$ exceeds two. This can be seen by noting that for each value of k, we randomly generated exactly 1000 sequences, and each row of the table sums

		$\hat{k}$		
		0	1	2
k	0	983	17	0
	1	686	312	2
	2	758	182	60

Table 2.3: The $(k, \hat{k})$ entry equals the number of times the BIC order estimator returned an order of $\hat{k}$, when applied to a random sequence of length $N = 500$ generated by a randomly generated Markov chain of true order k. For results in this table, transition probabilities were drawn uniformly from $(0.4, 0.6) \subset \mathbb{R}$.

to 1000. Based on the numbers given in Table 2.2, we make the following estimates:

$$P(k \geq 1 \,|\, \hat{k} = 1) \approx \frac{818 + 116}{33 + 818 + 116} = 0.9659 \tag{2.4}$$

$$P(k = 0 \,|\, \hat{k} = 0) \approx \frac{966}{966 + 180 + 28} = 0.8228 \tag{2.5}$$

That is to say, if the BIC order estimator equals one for a given sequence, we find there is a greater than 95% chance that the sequence was generated by a Markov chain of *at least* order one, i.e., a 95% chance that the sequence was in fact *not* IID. On the other hand, if the BIC order estimator equals zero for a given sequence, we find that there is an approximately 80% chance that the sequence was in fact IID.

In our observations, after converting real time series for stocks into u/d sequences, the maximum likelihood estimates of the transition probabilities are always between 0.4 and 0.6. This motivates us to rerun the above tests. This time, when we randomly generate transition probabilities for a k-th order Markov chain, we draw each probability uniformly from the interval $(0.4, 0.6)$. Other parameters of the test remain the same. The results are summarized in Table 2.3. Once again, in no instance do we find that the estimate $\hat{k}$ exceeds two; we ran 1000 tests for each value of k, and each row sums to 1000. Based on Table 2.3, we make the following estimates:

$$P(k \geq 1 \,|\, \hat{k} = 1) \approx \frac{312 + 182}{17 + 312 + 182} = 0.9667 \tag{2.6}$$

$$P(k = 0 \,|\, \hat{k} = 0) \approx \frac{983}{983 + 686 + 758} = 0.4050 \tag{2.7}$$

These results strengthen our conclusion that if the BIC order estimator applied to a sequence yields one, there is a greater than 95% chance that the sequence was in fact *not* IID. Note, however, that drawing the transition probabilities from the interval $(0.4, 0.6)$—centered at 0.5—has made it very easy for the BIC order estimator to *underestimate* the true order of the Markov chain. This is intuitively clear: if the transition probabilities for either a first- or second-order Markov chain are all close to 0.5, then short sequences generated by the Markov chain will appear to be IID. One will require an extremely long sequence from such a Markov chain in order to distinguish the sequence from an IID sequence; $N = 500$ samples is simply not enough.

The meaning of these results is that we can reliably use the BIC order estimator to *falsify* (2.1), but never to verify (2.1). In situations where we apply the BIC order estimator to real financial time series and obtain an estimate of at least one, there is a high probability that (2.1) is false; if, on the other hand, we obtain an estimate of zero, we should discard it.

2.5 Order Estimation: Results

We apply the BIC order estimation technique to stocks listed on the French CAC-40 index. Our interest in these stocks stems purely from the fact that European-style options on these stocks are traded on Euronext, and both the classical Black-Scholes model and our MT model are designed to price European-style options. For each stock on the index, we download at least two years of adjusted daily closing prices from Yahoo! Finance. Note that there are 252 trading days in one year, so at least two years' worth of data gives us a time series of length $N \geq 504$. We then apply the methodology of Section 2.4 and produce BIC order estimates for each time series. We find that there are six French companies for which the BIC order estimate equals one:

- Air Liquide (Euronext: AI), using data from Jan. 1, 2007 to Oct. 2, 2009.
- AXA Group (Euronext: CS, NYSE: AXA), using NYSE data from Feb. 1, 2007 to Oct. 2, 2009.
- L'Oréal Group (Euronext: OR), using data from Jan. 1, 2007 to Oct. 2, 2009.
- Pernod Ricard (Euronext: RI), using data from Jan. 1, 2003 to Oct. 2, 2009.
- Sanofi-Aventis (Euronext: SAN, NYSE: SNY), using either Euronext or NYSE data from June 30, 2007 to June 30, 2009.
- Société Générale (Euronext: GLE), using data from Jan. 1, 2007 to Oct. 2, 2009.

We believe that (2.1) is false for stock time series for each of these six companies. Later, when we compare the results of the Black-Scholes and MT models against market prices for European call options for these six companies, we will add further evidence to this claim. Note that we could also test (2.1) by using more traditional time series methods such as ACF and PACF. However,

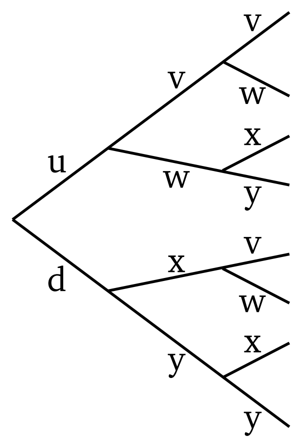

Figure 2.1: Illustration of the first three steps of the Markov tree. An upward edge always bifurcates into v and w. A downward edge always bifurcates into x and y. In this way, the tree accounts for the first-order Markov nature of the underlying asset's log return time series.

since our focus is obtaining a discrete tree model to price options, it seems natural to convert the original time series into a finite state time series and then test (2.1). In future work, we shall explore whether there exist time series whose non-IID behavior can be detected correctly by Markov order estimators and *not* by ACF-based methods, and vice versa.

2.6 Markov Tree Model: Theory

We now describe a tree model that accounts for the first-order Markov dependence in the log return time series. We restrict our model to accommodate only first-order Markov dependence (instead of, say, k-th order Markov dependence) not only to obtain computational tractability but also to maintain parsimony. Like the binomial tree, our tree is generated by working forward from valuation day to expiration of the option. Let S_n be the stock's spot price at time step n. When $n = 0$, we use one step of the standard binomial tree

$$P(S_1 = uS_0) = q \tag{2.8a}$$

$$P(S_1 = dS_0) = 1 - q. \tag{2.8b}$$

For $n \geq 1$, let us define two events:

$$S_n^+ = \{S_n \geq S_{n-1}\} \tag{2.9}$$

$$S_n^- = \{S_n < S_{n-1}\}. \tag{2.10}$$

In words, the event S_n^+ is the event that the stock price increased from time step $n-1$ to time step n. The event S_n^- is the complement of S_n^+, i.e., the event that the stock price decreased from time step $n-1$ to time step n. We can now write down our model for the evolution of S_n, for $n \geq 1$:

$$P(S_{n+1} = vS_n | S_n^+) = q^+ \tag{2.11a}$$

$$P(S_{n+1} = wS_n | S_n^+) = 1 - q^+ \tag{2.11b}$$

$$P(S_{n+1} = xS_n | S_n^-) = q^- \tag{2.11c}$$

$$P(S_{n+1} = yS_n | S_n^-) = 1 - q^-. \tag{2.11d}$$

Here we have introduced four symbols, v, w, x and y, which represent different factors by which the stock price at every time step is allowed to change. According to our model, if the stock price increased from step $n-1$ to step n, then the stock price at step $n+1$ is vS_n with probability q^+ and wS_n with probability $1-q^+$. If the stock price decreased from step $n-1$ to step n, then the stock price at step $n+1$ is xS_n with probability q^- and yS_n with probability $1-q^-$.

We remark that we think of q, q^+, and q^- as, respectively, risk-neutral versions of the empirical probabilities $P(u)$, $P(u|u)$, and $P(u|d)$. We shall explain later how, with respect to these risk-neutral probabilities, the stock price process S_n is in fact a martingale.

The first three steps of the tree are illustrated in Figure 2.1. If we let S_0 denote the initial spot price of the stock, then it is clear that $S_3 \in J_3$ where

$$J_3 = \{S_0uv^2, S_0uvw, S_0uwx, S_0uwy, S_0dxv, S_0dxw, S_0dyx, S_0dy^2\}.$$

In general, let J_n denote the vector of possible states the stock can be in after n steps of the Markov tree. Let $\delta_n : J_n \to \mathbb{Z}^+$ be the function that counts the number of paths in the tree that lead from S_0 to a given element of J_n. For $\omega \in J_n$, we refer to $\delta_n(\omega)$ as the *duplication number* of state ω. We list without proof these facts:

- J_n contains $n^2 - n + 2$ unique elements.

  That is to say, states do recombine. If the stock decreases from S_0uvw, it reaches the same value as if it increases from S_0uwx—in both cases, it reaches S_0uvwx. Because there are two possible paths leading from S_0 to S_0uvwx, we assign the duplication number $\delta_4(S_0uvwx) = 2$. Because states recombine, the number of states does *not* increase like 2^n. In the standard binomial model, the number of states grows linearly in the depth of the tree n. In the MT model, the number of states grows quadratically in the depth of the tree n. This polynomial growth ensures the tractability of the MT model as a computational method.

- $\sum_{\sigma \in J_n} \delta_n(\sigma) = 2^n$.

One can make sense of this intuitively by recalling that if we do not count the duplication of states, then a tree of depth n will contain 2^n states.

- Let p_n denote the polynomial that gives all states in J_n together with their duplication numbers, i.e.,

$$p_n(u,d,v,w,x,y) = \sum_{\omega \in J_n} \delta_n(\omega)\omega.$$

Then p_n may be computed via

$$p_n \quad = \quad u\begin{bmatrix} v & w & 0 & 0 \end{bmatrix} \begin{bmatrix} v & w & 0 & 0 \\ 0 & 0 & x & y \\ v & w & 0 & 0 \\ 0 & 0 & x & y \end{bmatrix}^{n-2} \mathbf{1} + d\begin{bmatrix} 0 & 0 & x & y \end{bmatrix} \begin{bmatrix} v & w & 0 & 0 \\ 0 & 0 & x & y \\ v & w & 0 & 0 \\ 0 & 0 & x & y \end{bmatrix}^{n-2} \mathbf{1},$$

where $\mathbf{1}$ denotes a column vector with each entry equal to one. The above fact may be derived by writing the adjacency matrix for a directed, weighted graph related to our Markov tree. For now, we merely mention that once we use the above iterative matrix formula to compute p_n for a given n and thereby generate a Markov tree of depth n, we can then reuse this tree many times to price many different options. For different options, S_0, u, d, v, w, x, and y will be different, but the set of states J_n and the duplication numbers δ_n will always be the same.

For example, carrying out the tree one step further than shown in Figure 2.1, we find that

$$J_4 = \{S_0uv^3, S_0uv^2w, S_0uvwx, S_0uvwy, S_0uw^2x, S_0uwyx, S_0uwy^2, \\ S_0dxv^2, S_0dxvw, S_0dx^2w, S_0dxwy, S_0dyxv, S_0dy^2x, S_0dy^3\}.$$

We have $\delta_4(S_0uvwx) = \delta_4(S_0dxwy) = 2$ and $\delta_4(\sigma) = 1$ for all other possible states $\sigma \in J_4$. Note that, as per our formula, there are $4^2 - 4 + 2 = 14$ elements in J_4, and $\sum_{\sigma \in J_4} \delta_4(\sigma) = 16 = 2^4$.

Next, note that it is simple to calculate the probability that the stock's price path reaches a given state in J_n, starting at S_0. Let $\sigma = S_0u^m d^{1-m} v^a w^b x^c y^d$ denote an arbitrary state in J_n. (Clearly either $m = 0$ or $m = 1$, and also the sum of the exponents must equal n, i.e., $1 + a + b + c + d = n$.). Then, by the definitions made in (2.8) and (2.11), the probability of reaching σ starting at S_0 is simply equal to

$$P(\sigma) = \delta_n(\sigma) q^m (1-q)^{1-m} \times \left(q^+\right)^a \left(1-q^+\right)^b \left(q^-\right)^c \left(1-q^-\right)^d. \tag{2.12}$$

Let us now explain how we use the tree to price a European call option. Let K denote the strike price and let S_* denote the spot price at the time of expiry. The payoff of the option is denoted by $(S_* - K)_+$, which equals zero unless $S_* - K > 0$, in which case it equals $S_* - K$. Let T denote the time until expiry. We fix the total number of steps N in the tree and set $\Delta t = T/N$. With these definitions, we can see that S_* is a random variable that can take on any of the states $\sigma \in J_N$ with probabilities given by (2.12). We have enough information to write down the expected value of the option's payoff at the time of expiry:

$$E[(S_* - K)_+] = \sum_{\sigma \in J_N} I_{\sigma > K}(\sigma - K)P(\sigma). \tag{2.13}$$

Here $I_{\sigma > K}$ is an indicator variable that equals one when $\sigma > K$ and zero when $\sigma \leq K$. Now let r equal the risk-free interest rate. Then we define the MT model's call option price to be expected payoff at the time of expiry, discounted to the present time:

$$C = e^{-rT} E[(S_* - K)_+]. \tag{2.14}$$

Note that using precisely the same approach, we can price European put options without making use of put-call parity. The payoff of a European put option equals $(K - S_*)_+$. The MT model's put option price is, once again, the discounted expected payoff:

$$U = e^{-rT} E[(K - S_*)_+]. \tag{2.15}$$

2.6.1 No Arbitrage.

Let us show that our model does not admit arbitrage. We define

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

One may easily verify that with these three risk-neutral probabilities,

$$E[S_1 | S_0] = uS_0 q + dS_0(1 - q) = e^{r\Delta t} S_0,$$

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 is enough to imply that the discounted stock process $\tilde{S}_n = e^{-rn\Delta t}S_n$ is a martingale under the risk-neutral probabilities given by q, q^+, and q^-. Then, by the first fundamental theorem of asset pricing (see (Shreve, 2004, Chapter 2.4)), there is no arbitrage in the MT model.

2.6.2 Implementation Notes.

The parameters u, d, v, w, x, and y are estimated as follows. For each date on which we wish to value an option, we start with the time series of one prior year's worth of adjusted closing daily returns for the stock. We scan through this time series and form two disjoint time series: each time a given day's return exceeds or equals the previous day's, we add that return to series 1; each time a given day's return is less than the previous day's, we add that return to series 2. We then take the logarithm of all returns in series 1 and 2 and also in the original time series. Let $\hat{\sigma}_+$ and l_+ denote the standard deviation and length of log return series 1, and let $\hat{\sigma}_-$ and l_- denote the standard deviation and length of log return series 2. Let $\hat{\sigma}$ be the standard deviation of the entire log return series. The standard deviations are then converted to volatilities σ, σ_+ and σ_- using $\sigma = \sqrt{252}\hat{\sigma}$ and $\sigma_\pm = \sqrt{l_\pm}\hat{\sigma}_\pm$. With these volatilities, we set

$$u = \exp\left(\sigma\sqrt{\Delta t}\right),$$
$$v = \exp\left(\sigma_+\sqrt{\Delta t}\right), \quad x = \exp\left(\sigma_-\sqrt{\Delta t}\right),$$

where Δt is the duration of each time step in the model. We then set $d = 1/u$, $w = 1/v$, and $y = 1/x$.

2.7 Tree Model: Results

For 44 trading days from July 17, 2009 to September 17, 2009, we tracked the end-of-day market prices for European-style call options for the six companies listed in Section 2.5. Data was obtained from euronext.com. We emphasize that all of the tests we are about to describe are out-of-sample tests; at no time did we use past or present market prices of options as inputs to the MT or Black-

Scholes models. The fact that the MT model requires no calibration with options price data from real markets is in marked contrast to, say, Rubinstein's implied binomial tree model (Rubinstein, 2012).

On day i of the study, we used stock and futures prices from days before or on day i to estimate parameters that are fed as inputs to the MT and Black-Scholes options pricing models. Specifically, we estimated the risk-free rate r and the volatilities σ, σ_+, and σ_-, which determine the jumps u, d, v, w, x, and y. With these parameters, we priced all exchange-traded options using both the MT and Black-Scholes models. For the MT model, we used $N = 501$ steps. For each option at hand, we compared the outputs of these options pricing models on day i to the market price of the same option on day i.

2.7.1 Comparison of Model and Market Prices.

We first consider the day-by-day performance of the MT model versus the Black-Scholes model, averaged across all strikes. Let i be a fixed day. Let $\mathbf{b}_i$, $\mathbf{m}_i$, and $\mathbf{M}_i$ be the vectors containing Black-Scholes, MT, and market prices on day i for options of different strikes (but the same expiration date). On each day, we compute

$$\epsilon_i^b = \frac{\|\mathbf{b}_i - \mathbf{M}_i\|_2}{\|\mathbf{M}_i\|_2}, \quad \epsilon_i^m = \frac{\|\mathbf{m}_i - \mathbf{M}_i\|_2}{\|\mathbf{M}_i\|_2}. \tag{2.16}$$

In each of the six panels of Figure 2.2, we plot the relative error curves ϵ_i^b (in red) and ϵ_i^m (in blue) versus day i for options from each of the six companies listed in Section 2.5, respectively.

We then consider the strike-by-strike performance of the MT model versus the Black-Scholes model, averaged across all days. Let j be a fixed strike price. Let $\mathbf{b}_j$, $\mathbf{m}_j$, and $\mathbf{M}_j$ be the vectors containing Black-Scholes, MT, and market prices for options with strike j on different days (but the same expiration date). On each day, we compute

$$\gamma_j^b = \frac{\|\mathbf{b}_j - \mathbf{M}_j\|_2}{\|\mathbf{M}_j\|_2}, \quad \gamma_j^m = \frac{\|\mathbf{m}_j - \mathbf{M}_j\|_2}{\|\mathbf{M}_j\|_2}. \tag{2.17}$$

In each of the six panels of Figure 2.3, we plot the log relative error curves $\log(\gamma_j^b)$ (in red) and $\log(\gamma_j^m)$ (in blue) versus strike price j for options from each of the six companies listed in Section 2.5, respectively.

In both Figure 2.2 and Figure 2.3, the following symbols are used to denote common expiration dates: "$\circ$" means September 2009, "$*$" means March 2010, "$\diamond$" means September 2010, and "$+$" means March 2011.

Comparing Black-Scholes and MT relative errors for options with the *same* expiration date means comparing blue and red curves with *identical* symbols in Figure 2.2 and Figure 2.3. For example, in Figure 2.2, comparing blue and red curves with "$+$" symbols shows that the MT model's prices for options expiring in March 2011 are much closer to market prices than the Black-Scholes model's prices for options expiring in March 2011. This is true for all six companies.

In fact, comparing blue and red curves in Figure 2.2 with identical symbols reveals that the only expiration date for which the two models produce comparable results is the September 2009 expiration date, denoted by "$\circ$." In this case, for all six companies, the MT model still produces relative errors ϵ_i^m that are two to ten times smaller than the relative errors ϵ_i^b produced by the Black-Scholes model. For all other expiration dates, comparing the two models on a day-by-day basis, the MT model's call option prices are far closer to market prices than the Black-Scholes model's prices.

Moving to Figure 2.3, we see that as the strike price increases, the Black-Scholes model's error increases more rapidly than the MT model's error. Note that each point on each of the panels in Figure 2.3 is an aggregate result, averaged (in the sense of the 2-norm) over 44 trading days' worth of data. For this reason, we believe Figure 2.3 provides strong evidence that the discrepancy between Black-Scholes and market prices for out-of-the-money options is not entirely due to the dependence of volatility on strike price and time until expiration.

2.7.2 Comparison of Volatilities.

For each of the six stocks listed in Section 2.5, we plot in Figure 2.4 the three volatilities σ, σ_+, and σ_- on each of the 44 days. These plots show that for three of the six stocks (AI, OR, and GLE before day 40), the difference between σ_+ and σ_- is small, on the order of 1%. For these three stocks, the MT model, in the way we have implemented it with the formulas from Section 4.3, produces option prices close to those produced by a binomial model with with volatility given by either σ_+, σ_-, or perhaps a weighted average of these values. It is noteworthy that a binomial model with *volatility estimated by splitting past historical data based on whether returns were increasing or decreasing relative to the previous day* does far better at tracking market prices than a vanilla Black-Scholes (or, equivalently, binomial) model with volatility σ. The formulas given in Section 4.3 for $\sigma_\pm$ were determined by extensive trial-and-error. In future work, we shall provide a more rigorous theory for estimating the parameters v, w, x, and y that serve as inputs to the MT model.

On the other hand, the plots in Figure 2.4 also indicate that for three of the six stocks (CS, RI, and SAN), the difference between σ_+ and σ_- is closer to 10%. In this case, one can show that the set of states J_{501} together with (2.12) yield a probability distribution on the set of stock prices at the time of expiry that is different from the distribution of final stock prices provided by a standard binomial model. For these three stocks, the MT model does not reduce to a binomial model. Depending on the specific values of the parameters, it is possible for the MT model's final stock price distribution to feature heavier tails and interesting asymmetries relative to the lognormal distribution. We expect that these features, which have been reported elsewhere in the financial time series literature, will appear when we use better methods for estimating $\sigma_{\pm}$.

2.8 Conclusion

Over the past two years, many securities have been subject to large fluctations in price, and financial modeling assumptions that used to be considered standard should now be called into question. One such assumption is (2.1). In this Chapter, we have tested (2.1) using the BIC order estimation method. The tests have revealed six stocks in the French CAC-40 index whose log return time series is not IID. For these six stocks, and for other stocks whose log return time series is best modeled by a k-th order Markov chain with $k \geq 1$, we propose the MT options pricing model. The number of states in the Markov tree grows quadratically in the depth of the tree, giving the model computational tractability. Implementing the MT model, we find strong agreement between the MT model's prices and market prices.

In future work, we shall compare the MT model against more sophisticated options pricing models, such as those incorporating stochastic volatility. The first-order Markov dependence of our tree model is a general concept that could be incorporated into discrete-time stochastic volatility models (Florescu and Viens, 2008), which could further reduce the error between model and market prices. Finally, we shall extend the MT model to price weather derivatives, especially in light of scientific studies that propose Markov chain models for quantities such as rainfall (Gabriel and Neumann, 1962).

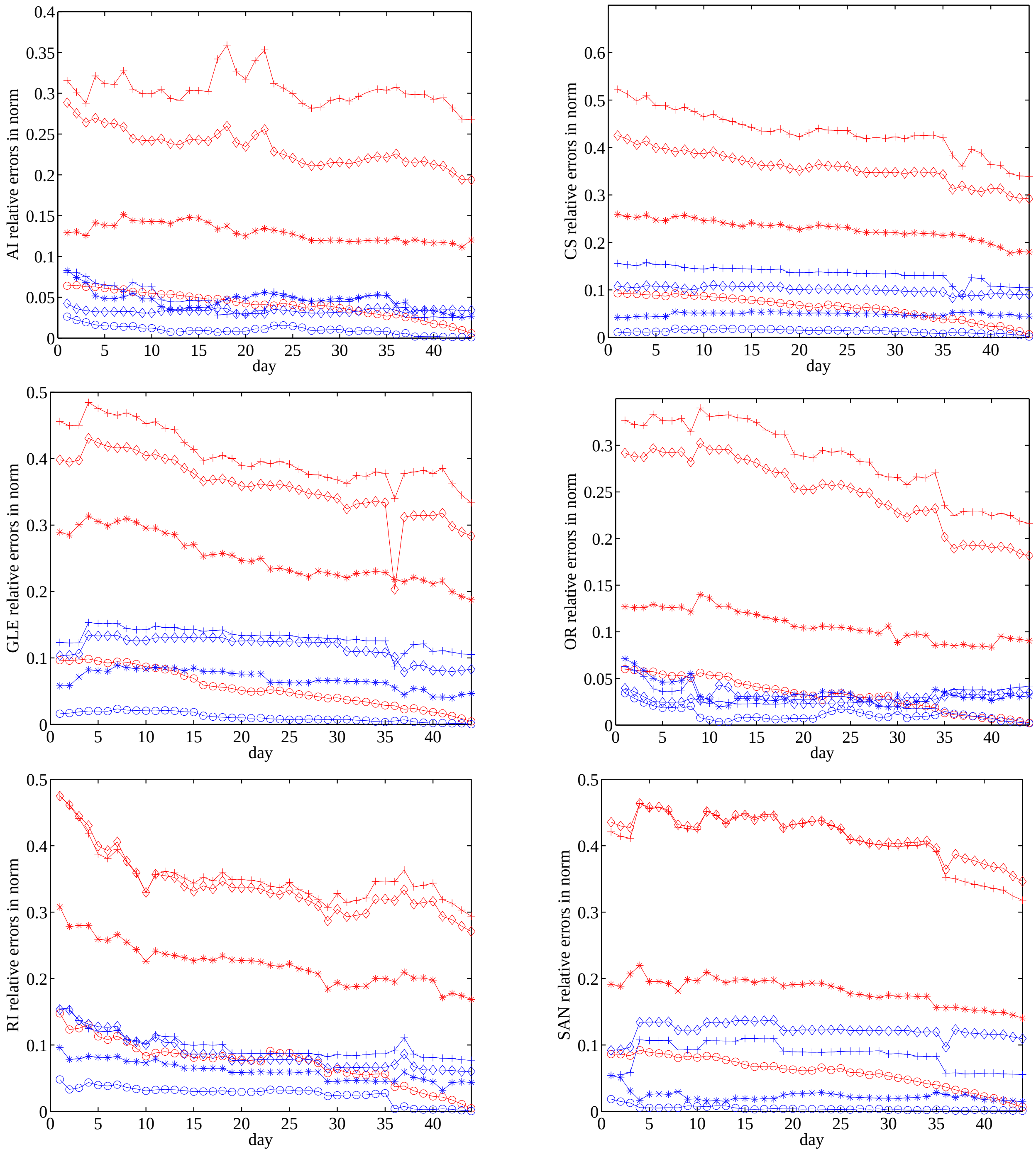

Figure 2.2: From left to right, top to bottom, we plot model relative errors for the six companies listed in Section 2.5 in the following order (alphabetical in the Euronext symbols): AI, CS, GLE, OR, RI, and SAN. Each panel displays relative errors ϵ_i^b (Black-Scholes error in red) and ϵ_i^m (MT error in blue) versus day i for options with different expiration dates. The following symbols are used to denote common expiration dates: "$\circ$" means September 2009, "$*$" means March 2010, "$\diamond$" means September 2010, and "$+$" means March 2011. *Note that for all expiration dates except September 2009, the MT model's relative error curves are far below the Black-Scholes relative error curves.* For options expiring in September 2009, both models yield nearly identical results.

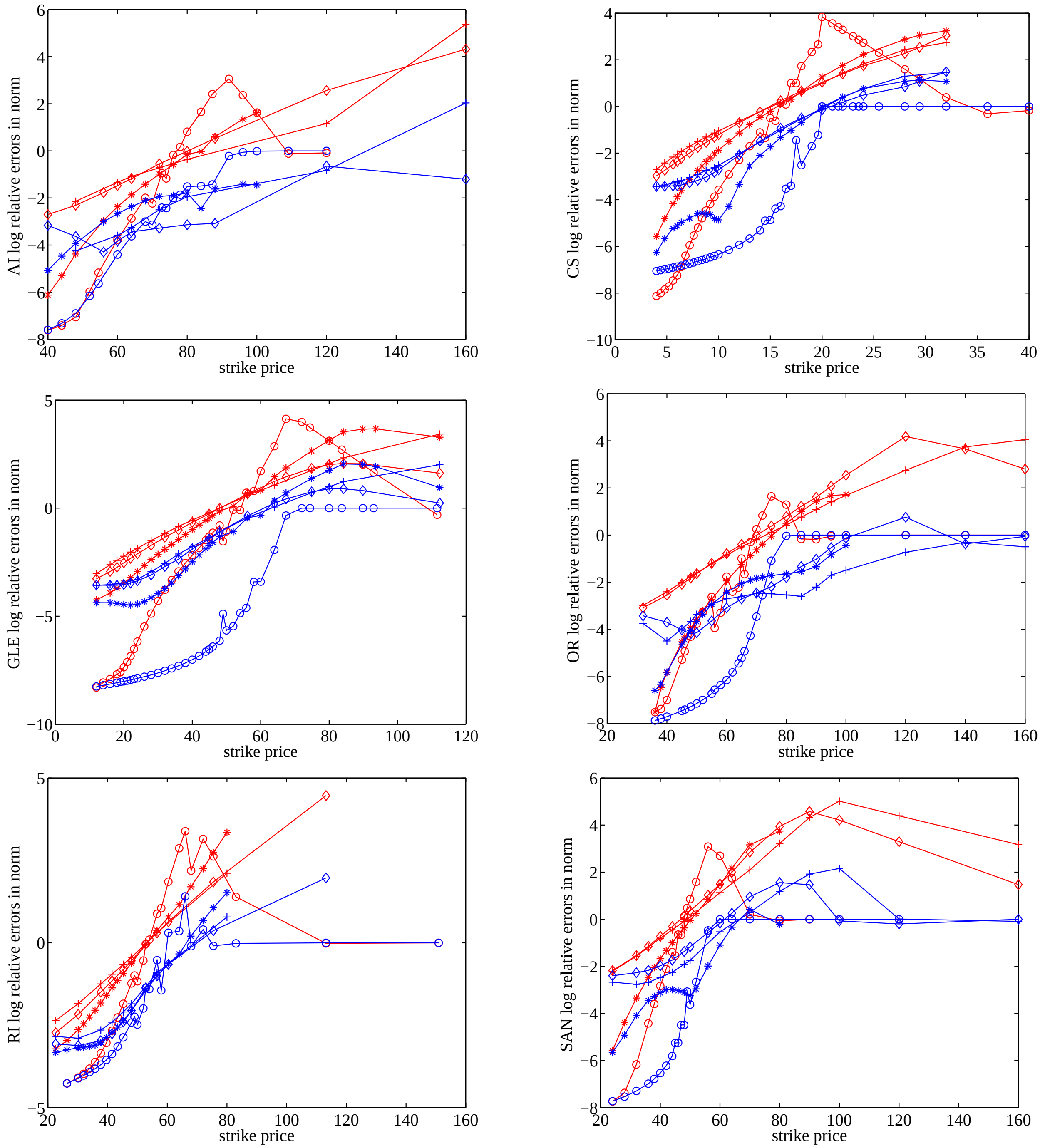

Figure 2.3: From left to right, top to bottom, we plot model relative errors for the six companies listed in Section 2.5 in the following order (alphabetical in the Euronext symbols): AI, CS, GLE, OR, RI, and SAN. Each panel displays log relative errors $\log(\gamma_j^b)$ (Black-Scholes error in red) and $\log(\gamma_j^m)$ (MT error in blue) versus strike price j for options with different expiration dates. The following symbols are used to denote common expiration dates: "$\circ$" means September 2009, "$*$" means March 2010, "$\diamond$" means September 2010, and "$+$" means March 2011. *Note that for all expiration dates, as the strike price increases, the Black-Scholes model's relative error curves far exceed the MT model's relative error curves.*

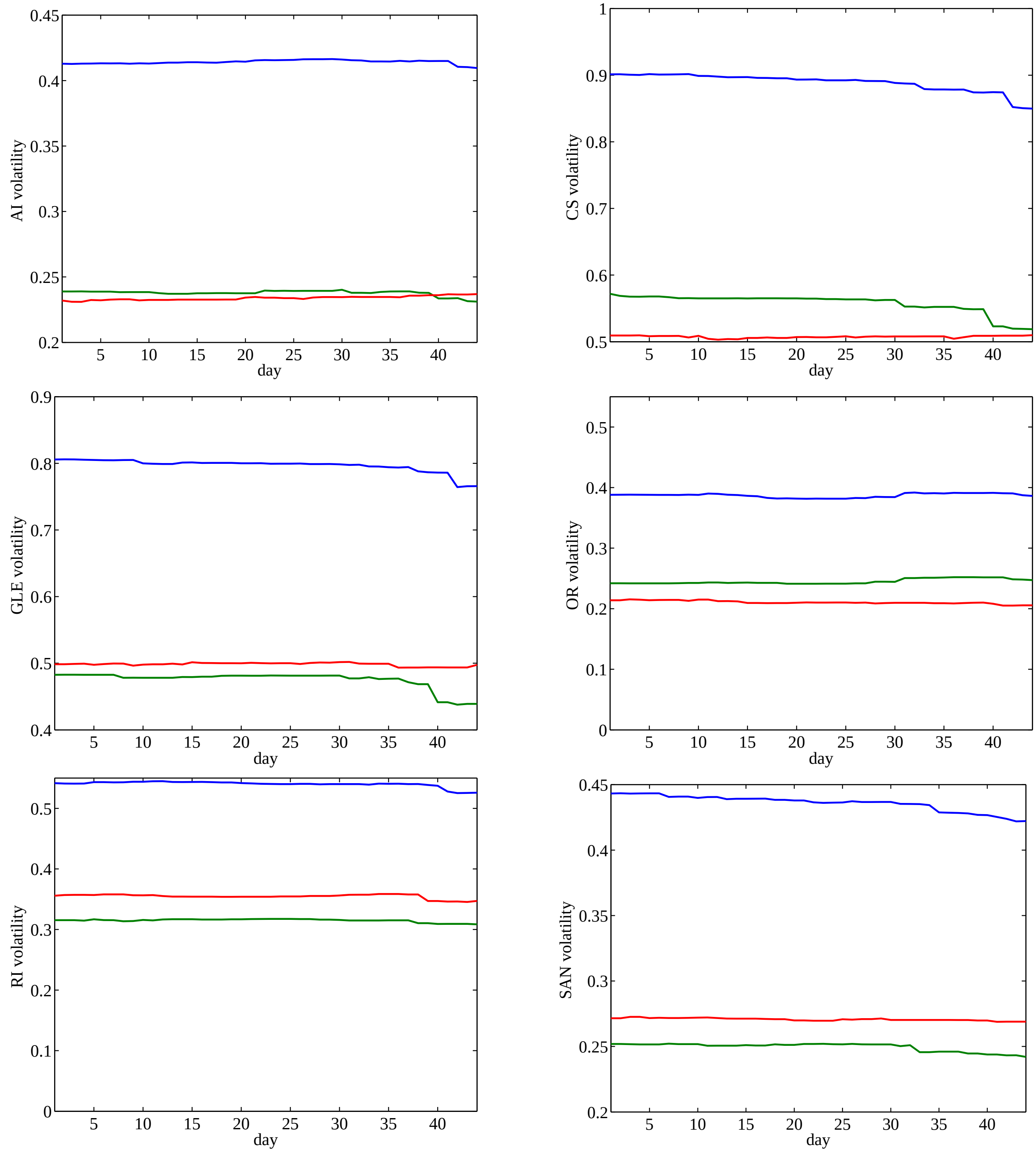

Figure 2.4: From left to right, top to bottom, we plot the volatilities for six companies listed in Section 2.5 in the following order (alphabetical in the Euronext symbols): AI, CS, GLE, OR, RI, and SAN. Each panel displays the volatilies σ, σ_+ and σ_- in blue, green and red respectively versus day i. These values σ, σ_+, and σ_- are used to calculate the jump factors u, v and x, respectively. Recall that the jump factors d, w, and y are the reciprocals of u, v, and x, respectively. *Note that in the MT model, volatility is assumed to be independent of the strike price and the expiration date of the option.*

Chapter 3

Markov Tree: Continuous Model

3.1 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 Chapter 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. 3.1. 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 Chapter, 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.

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) (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; Appleby et al., 2012a,b). 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 Chapter is as follows. In Section 3.2, 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 3.3, 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 3.4, we use the approximate continuous p.d.f. to derive a closed-form option price. In Section 3.5, 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.

3.2 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.

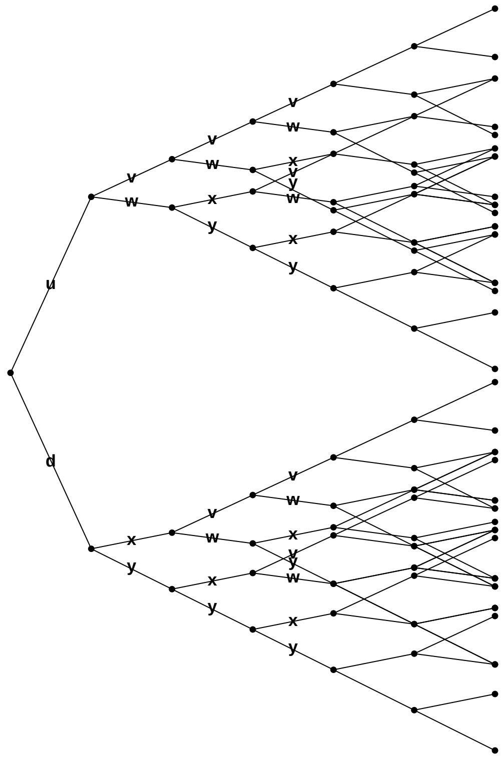

Figure 3.1: Tree of depth $n = 6$ showing recombination of paths of the underlying asset in the MT model. The asset begins with price S_0 and is multiplied by the weights along the path. For example, a possible path of length 3 shown here is S_0uwx. Both the probabilities and the outcomes of S_{n+1}/S_n 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.

3.2.1 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:

$$\tilde{S}_{n+1}(\omega) = \tilde{S}_n(\omega) + G(\mathcal{H}(\tilde{S}_n - \tilde{S}_{n-1}), \omega), \tag{3.1}$$

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

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

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

$$P(G(1,\omega) = \log u) = q$$
$$P(G(1,\omega) = \log d) = 1 - q.$$

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

3.2.2 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

$$n_{HH} + n_{HT} + n_{TH} + n_{TT} = n - 1. \tag{3.2}$$

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*:

$$t = n_{HT} + n_{TH}. \tag{3.3}$$

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 (3.2).

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.

3.2.3 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

$$\#(\mathbf{v}) = \binom{n_{HH}+a-1}{a-1}\binom{n_{TT}+b-1}{b-1} = \binom{n_{HH}+n_{TH}}{n_{TH}}\binom{n_{TT}+n_{HT}-1}{n_{HT}-1}. \tag{3.4}$$

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

$$\#(\mathbf{v}) = \binom{n_{HH}+a-1}{a-1}\binom{n_{TT}+b-1}{b-1} = \binom{n_{HH}+n_{TH}-1}{n_{TH}-1}\binom{n_{TT}+n_{HT}}{n_{HT}} \tag{3.5}$$

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

$$P(\tilde{S}_n = \tilde{S}_0 + \mathbf{s}\cdot\mathbf{v}) = \#(\mathbf{v})\,\mathbf{q}^{\mathbf{v}}, \tag{3.6}$$

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.

3.3 Continuous Approximation of the Markov Tree

We can see from (2.13) 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. (3.6) generated by the tree. Though we have developed an efficient algorithm to generate all states of the tree, the quantity $\#(\mathbf{v})$ defined by (3.4) and (3.5) 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.

3.3.1 Recursion

To develop a continuous approximation, we first rewrite the discrete-time process (3.1) as a recursion. We assume all movements are symmetric about one (i.e., $d = 1/u$, $w = 1/v$, $y = 1/x$) and define

$$l_u = \log u = -\log d \tag{3.7a}$$

$$l_1 = \log v = -\log w \tag{3.7b}$$

$$l_2 = \log x = -\log y \tag{3.7c}$$

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

$$R(n, \tilde{s}) = q^+ R(n-1, \tilde{s} - l_1) + q^- L(n-1, \tilde{s} - l_2). \tag{3.8}$$

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

$$L(n, \tilde{s}) = (1-q^+)R(n-1, \tilde{s}+l_1) + (1-q^-)L(n-1, \tilde{s}+l_2). \tag{3.9}$$

3.3.2 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 $\tilde{s}$:

$$\hat{f}(k) = \int_{\mathbb{R}} f(\tilde{s})e^{-ik\tilde{s}}\, d\tilde{s}, \qquad f(\tilde{s}) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(k)e^{ik\tilde{s}}\, dk. \tag{3.10}$$

Define

$$M = \begin{bmatrix} q^+e^{-ikl_1} & q^-e^{-ikl_2} \\ (1-q^+)e^{ikl_1} & (1-q^-)e^{ikl_2} \end{bmatrix}. \tag{3.11}$$

Then, taking the Fourier transforms of both sides of (3.8) and (3.9), we are able to put the system into matrix-vector form and solve:

$$\begin{bmatrix} \hat{R}(n,k) \\ \hat{L}(n,k) \end{bmatrix} = M \begin{bmatrix} \hat{R}(n-1,k) \\ \hat{L}(n-1,k) \end{bmatrix} = M^{n-1} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix}. \tag{3.12}$$

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 (3.12) with the row vector $\mathbf{1}^t$:

$$\hat{P}(n,k) = \mathbf{1}^t M^{n-1} \begin{bmatrix} \hat{R}(1,k) \\ \hat{L}(1,k) \end{bmatrix}. \tag{3.13}$$

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).

3.3.3 Numerical solution in real space

In the numerical inversion of (3.13), 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. 3.2. 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 (3.13) 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 (3.8) and (3.9) about l_1 and l_2 and then numerically solving the partial differential equation thus obtained.

Note that even though numerical Fourier inversion of (3.13) yields a fast, accurate approx-

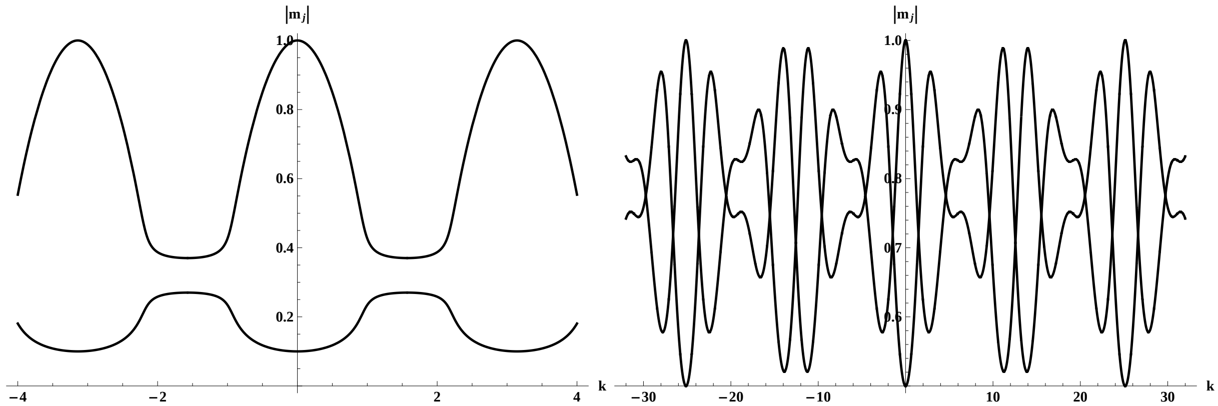

Figure 3.2: Moduli of the eigenvalues m_1, m_2 of the matrix M defined in (3.11). We plot $|m_j|$ as a function of Fourier variable k to show that, for almost all values of k, the eigenvalues are in the interior of the unit disc in $\mathbb{C}$. For the plot on the left, we set $l_1 = l_2 = 1$, $q^+ = 3/5$, $q^- = 1/2$. 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.

imation 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 (3.13) only to compare against the true Markov tree p.m.f. (3.6) and the asymptotic approximation that we derive next.

3.3.4 Asymptotic solution in real space

We now derive an asymptotic approximation to the p.d.f. (Rudnick and Gaspari, 2004, Chap. 5.2) that uses generating functions. 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 (3.12), we can write

$$\begin{aligned}\begin{bmatrix}\rho(z,k)\\ \lambda(z,k)\end{bmatrix} &= \sum_{n=0}^{\infty} M^n z^n \begin{bmatrix}\hat{R}(1,k)\\ \hat{L}(1,k)\end{bmatrix}\\ &= (\mathrm{I} - Mz)^{-1}\begin{bmatrix}\hat{R}(1,k)\\ \hat{L}(1,k)\end{bmatrix}\\ &= \frac{1}{(1-zm_1)(1-zm_2)}\begin{bmatrix}1-(1-q^-)e^{ikl_2}z & q^- e^{-ikl_2}z\\ (1-q^+)e^{ikl_1}z & 1-q^+e^{-ikl_1}z\end{bmatrix}\begin{bmatrix}\hat{R}(1,k)\\ \hat{L}(1,k)\end{bmatrix}.\end{aligned} \tag{3.14}$$

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

$$\begin{aligned} p(z,k) &= \sum_{n=0}^{\infty} \hat{P}(n+1,k) z^n \\ &= \sum_{n=0}^{\infty} \left(\hat{R}(n+1,k) + \hat{L}(n+1,k) \right) z^n \\ &= \mathbf{1}^t \begin{bmatrix} \rho(z,k) \\ \lambda(z,k) \end{bmatrix}. \end{aligned} \tag{3.15}$$

Substituting (3.14) in (3.15) 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 $p(z,k)$

$$\begin{aligned} &= \frac{\hat{P}(1,k)}{(1-zm_1)(1-zm_2)} + z\frac{\gamma}{(1-zm_1)(1-zm_2)} \\ &= \frac{\hat{P}(1,k)}{m_1-m_2} \left(\frac{m_1}{1-zm_1} - \frac{m_2}{1-zm_2} \right) + z\frac{\gamma}{m_1-m_2} \left(\frac{m_1}{1-zm_1} - \frac{m_2}{1-zm_2} \right) \\ &= \frac{\hat{P}(1,k)}{m_1-m_2} \left(m_1 \sum_{n=0}^{\infty} m_1^n z^n - m_2 \sum_{n=0}^{\infty} m_2^n z^n \right) + z\frac{\gamma}{m_1-m_2} \left(m_1 \sum_{n=0}^{\infty} m_1^n z^n - m_2 \sum_{n=0}^{\infty} m_2^n z^n \right) \\ &= \frac{1}{m_1-m_2} \sum_{n=0}^{\infty} \left(\hat{P}(1,k)(m_1^{n+1} - m_2^{n+1}) + \gamma(m_1^n - m_2^n) \right) z^n. \end{aligned} \tag{3.16}$$

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 (3.16):

$$\hat{P}(n+1,k) = \frac{1}{m_1-m_2} \left(\hat{P}(1,k)(m_1^{n+1} - m_2^{n+1}) + \gamma(m_1^n - m_2^n) \right). \tag{3.17}$$

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 (3.13).

Since m_1, m_2, and γ all depend on k, we cannot expect to find a closed-form inverse Fourier transform of (3.17). 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:

$$m_1 = 1 + i\zeta_{11}k - \zeta_{12}k^2 + O(k^3) \tag{3.18}$$
$$m_2 = (q^+ - q^-) + i\zeta_{21}k - \zeta_{22}k^2 + O(k^3) \tag{3.19}$$

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 $|m_{1,2}| \leq 1$, when n is large, we get

$$\begin{aligned}\hat{P}(n+1,k) &\sim \frac{1}{m_1 - m_2}\left(\hat{P}(1,k)m_1^{n+1} + \gamma m_1^n\right)\\ &= \frac{1}{m_1-m_2}\left(qF_1(m_1^{n+1} + \alpha m_1^n) + (1-q)F_2(m_1^{n+1} + \beta m_1^n)\right)\\ &\sim m_1^{n-1}\left(qF_1(m_1+\alpha) + (1-q)F_2(m_1+\beta)\right).\end{aligned} \tag{3.20}$$

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

$$\begin{aligned}m_1^{n-1}qF_1(m_1+\alpha) &= \exp\left(\log q + \log(F_1m_1 + F_1\alpha) + (n-1)\log m_1\right)\\ &= q\exp\left(\log(F_1m_1 + F_1\alpha) + (n-1)\log m_1\right) \qquad (3.21)\\ &\sim q\exp\left(-\mu_1 ik - \frac{\sigma_1^2}{2}k^2\right). \qquad (3.22)\end{aligned}$$

To pass from (3.21) to (3.22), 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 B.1. Proceeding analogously for the second term in (3.20), 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 B.1. We now express $\hat{P}(n,k)$ as

$$\hat{P}(n+1,k) \sim q\exp\left(-\mu_1 ik - \frac{\sigma_1^2}{2}k^2\right) + (1-q)\exp\left(-\mu_2 ik - \frac{\sigma_2^2}{2}k^2\right). \tag{3.23}$$

Taking the inverse Fourier transform of both sides of (3.23), we obtain the approximate p.d.f.

$$P(n+1,\tilde{s}) \sim f_{\tilde{s}}(\tilde{s},n+1) := \frac{q}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(\tilde{s}-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1-q}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(\tilde{s}-\mu_2)^2}{2\sigma_2^2}\right). \quad (3.24)$$

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

$$f_{\tilde{s}}(\tilde{s},n+1) = qg_1(\tilde{s},n+1) + (1-q)g_2(\tilde{s},n+1). \quad (3.25)$$

3.3.5 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. 3.3, we plot the c.d.f.'s obtained from the following probability mass/density functions: MT, the exact Markov tree p.m.f. (3.6), FT, the p.d.f. obtained by numerical inversion of the Fourier transform (3.13), and Asym, the p.d.f. (3.24) obtained by asymptotic approximation. Table 3.1 shows the parameters used in the comparison in each of the panels.

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 3.1 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 3.3.3 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 (3.24) and (3.25) to price options.

3.4 Option Price

Pricing a European call option using the normal mixture distribution (3.24) and (3.25) 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 (2.13) and evaluate the expected value using the p.d.f. (3.24):

$$C = e^{-rY} \int_K^\infty (s-K) f_s(s,Y)\, ds, \quad (3.26)$$

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 $N \gg 100$.

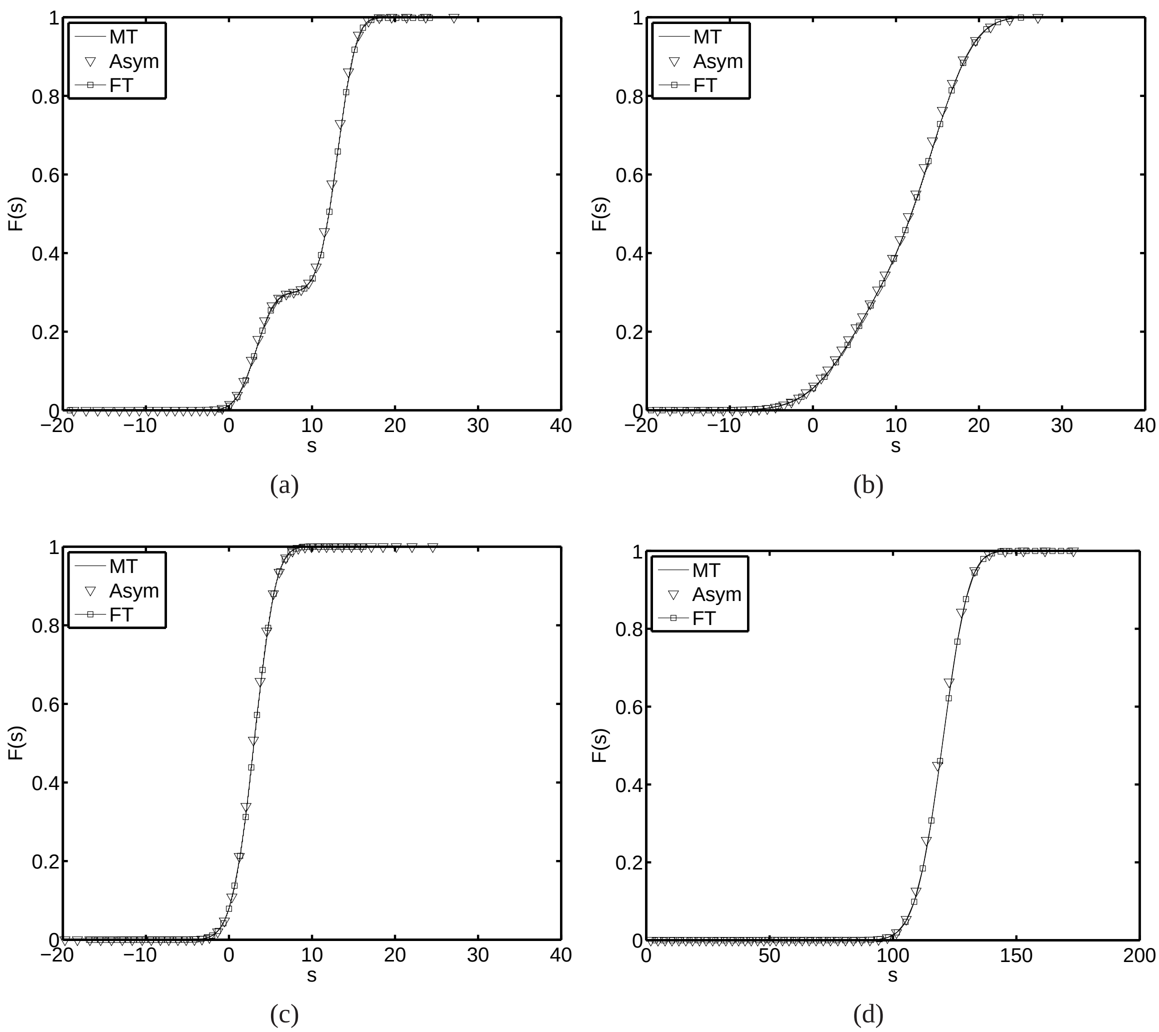

Figure 3.3: Comparison of cumulative distribution functions for MT, the exact Markov tree p.m.f. (3.6), FT, the p.d.f. obtained by numerical inversion of the Fourier transform (3.13), and Asym, the p.d.f. (3.24) obtained by asymptotic approximation. Table 3.1 shows the parameters used in the comparison in each of the panels.

Panel of Fig. 3.3	l_u	l_1	l_2	q	q^+	q^-	N	$\|\mathrm{FT}-\mathrm{MT}\|_\infty$	$\|\mathrm{Asym}-\mathrm{MT}\|_\infty$
3.3a	5.0	0.2	0.3	0.7	0.4	0.8	150	0.0097	0.0362
3.3b	5.0	0.2	0.3	0.7	0.8	0.4	150	0.0042	0.0247
3.3c	0.05	0.2	0.3	0.5	0.3	0.7	150	0.0117	0.0320
3.3d	0.05	0.4	0.6	0.5	0.8	0.7	500	0.0065	0.0403

Table 3.1: Details of parameters used for each panel in Fig. 3.3 and numerical values of the errors.

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 (3.26) and use the decomposition (3.25)

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

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

$$Ce^{rY} = qS_0 \exp\left(\frac{\sigma_1^2}{2} + \mu_1\right)\Phi(x_1) + (1-q)S_0 \exp\left(\frac{\sigma_2^2}{2} + \mu_2\right)\Phi(x_2) - qK\Phi(x_3) - (1-q)K\Phi(x_4), \tag{3.27}$$

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

$$x_i = \frac{\mu_i + \sigma_i^2 + \log(S_0/K)}{\sigma_i}, \quad x_{i+2} = \frac{\mu_i + \log(S_0/K)}{\sigma_i} \tag{3.28}$$

for $i \in \{1, 2\}$. 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).

3.5 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.

3.5.1 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

$$l_u = \sigma\sqrt{\Delta t}, \quad l_1 = \sigma^+\sqrt{\Delta t}, \quad l_2 = \sigma^-\sqrt{\Delta t}. \tag{3.29}$$

Then u, d, v, w, x, and y are defined by (3.7), enabling us to calculate the risk-neutral probabilities via (2.6.1), the mixture parameters (μ_j, σ_j) defined in the Appendix B.1, and the call option price defined by (3.27).

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

$$\hat{\sigma}^+ = \kappa \,\mathrm{mean} \left| Z^+ - \mathrm{mean}\left(Z^+\right)\right|, \qquad \hat{\sigma}^- = \kappa \,\mathrm{mean} \left| Z^- - \mathrm{mean}\left(Z^-\right)\right|. \tag{3.30}$$

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 Chapter, 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 (3.30). 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 $\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^{+}$, and $\hat{\boldsymbol{\sigma}}^{-}$ 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 $\hat{\boldsymbol{\sigma}}$:

$$\boldsymbol{\sigma}_{*}^{+} = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} \end{bmatrix} \boldsymbol{\eta}_{\mathbf{1}}^{+} + \boldsymbol{\varepsilon}_{\mathbf{1}}^{+} \tag{3.31a}$$
$$\boldsymbol{\sigma}_{*}^{-} = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} \end{bmatrix} \boldsymbol{\eta}_{\mathbf{1}}^{-} + \boldsymbol{\varepsilon}_{\mathbf{1}}^{-} \tag{3.31b}$$

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 3.2.

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

$$\boldsymbol{\sigma}_{*}^{+} = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} & \hat{\boldsymbol{\sigma}}^{+} & \hat{\boldsymbol{\sigma}}^{-} \end{bmatrix} \boldsymbol{\eta}_{\mathbf{3}}^{+} + \boldsymbol{\varepsilon}_{\mathbf{3}}^{+} \tag{3.32a}$$
$$\boldsymbol{\sigma}_{*}^{-} = \begin{bmatrix} \mathbf{1} & \hat{\boldsymbol{\sigma}} & \hat{\boldsymbol{\sigma}}^{+} & \hat{\boldsymbol{\sigma}}^{-} \end{bmatrix} \boldsymbol{\eta}_{\mathbf{3}}^{-} + \boldsymbol{\varepsilon}_{\mathbf{3}}^{-} \tag{3.32b}$$

The adjusted R^2 values for this model can be found in the "Three parameters" L^2 columns of Table 3.2.

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

$$\boldsymbol{\sigma}_{*}^{+} = \psi_1^{+}(\hat{\boldsymbol{\sigma}}) + \boldsymbol{\delta}_{\mathbf{1}}^{+} \tag{3.33a}$$
$$\boldsymbol{\sigma}_{*}^{-} = \psi_1^{-}(\hat{\boldsymbol{\sigma}}) + \boldsymbol{\delta}_{\mathbf{1}}^{-} \tag{3.33b}$$

The second model can be written

$$\boldsymbol{\sigma}_{*}^{+} = \psi_3^{+}(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^{+}, \hat{\boldsymbol{\sigma}}^{-}) + \boldsymbol{\delta}_{\mathbf{3}}^{+} \tag{3.34a}$$
$$\boldsymbol{\sigma}_{*}^{-} = \psi_3^{-}(\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\sigma}}^{+}, \hat{\boldsymbol{\sigma}}^{-}) + \boldsymbol{\delta}_{\mathbf{3}}^{-} \tag{3.34b}$$

The adjusted R^2 values for models (3.33) and (3.34) can be found in Table 3.2, 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.

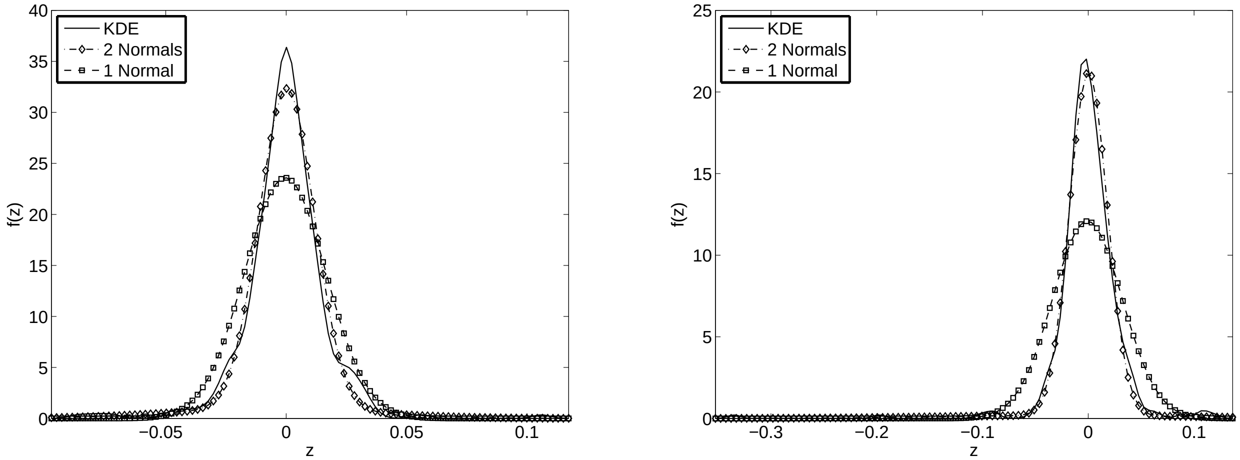

Figure 3.4: We fit three p.d.f.'s to daily log return time series for GOOG (left) and DF (right). The p.d.f's are a kernel density estimate (KDE), a mixture of two normals, and a single normal. The results show that the mixture of two normals more closely matches the KDE density. See Section 3.5.2 for more details.

In the remainder of this Chapter, 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 (3.34). 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.

3.5.2 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. 3.4, 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.

	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 3.2: Adjusted R^2 values for the linear models (3.31) and (3.32) are given in the L^2 subcolumns with respective column headings "One parameter" and "Three parameters." Adjusted R^2 values for the tree models (3.33) and (3.34) 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.

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 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.

3.5.3 Comparing model and market option prices

We now test the models MT Naive and MT Reg, introduced in Section 4.3, 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

	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 3.3: 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 3.4: 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 3.3.

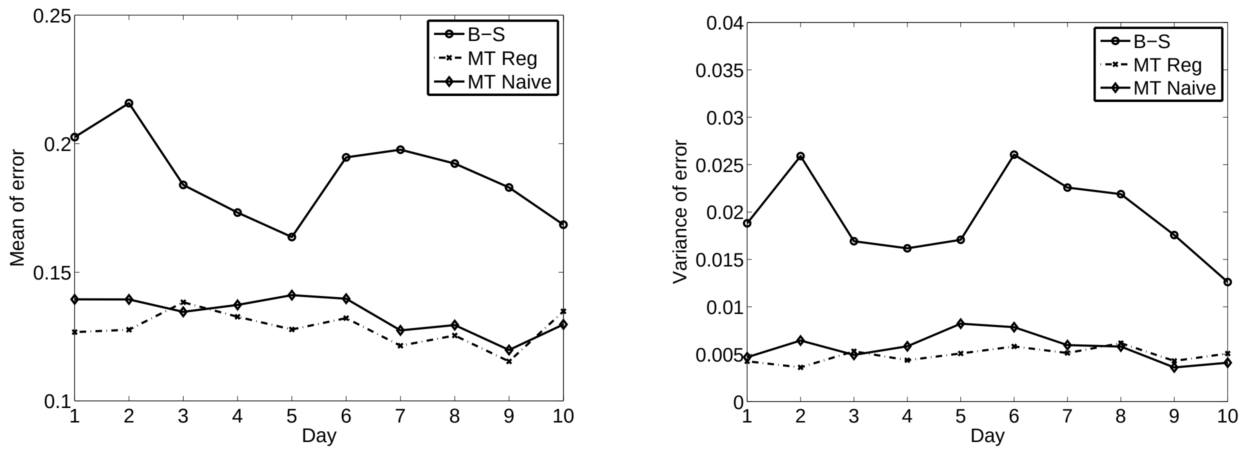

Figure 3.5: The left (respectively, right) panel shows E^{model} (respectively, $\text{Var}[E^{\text{model}}(\theta)]$) for each of 10 days of testing and each of the three models B-S, MT Reg, and MT Naive. See Section 3.5.3 for more details.

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 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. 3.5, 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 3.3 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 3.5.2 for more details.

In the right panel of Fig. 3.5, 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 3.4 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. 3.5 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 3.3 and 3.4 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. 3.6. 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. 3.7. Again we have 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Naive}}(\theta))$. 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.

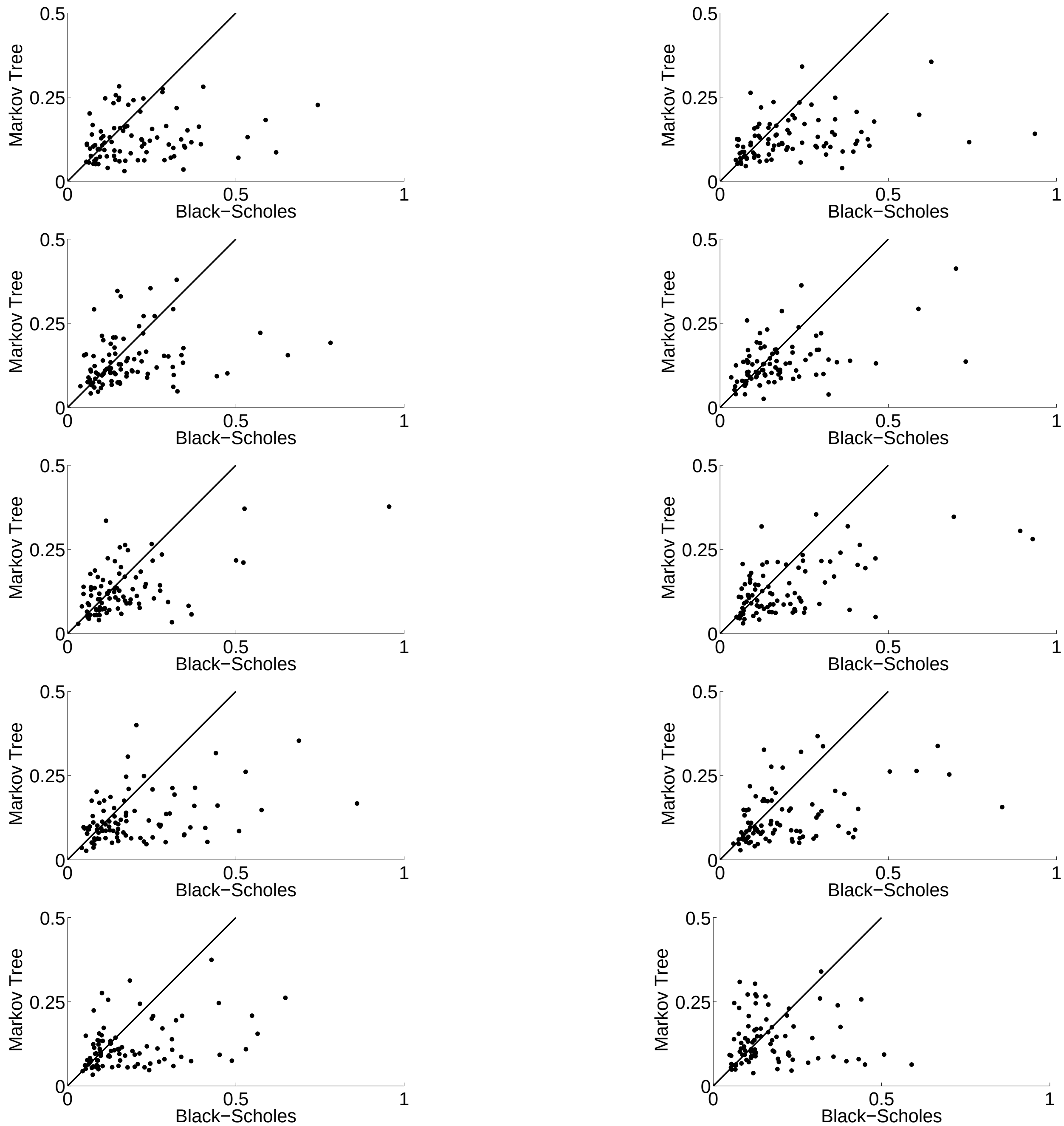

Figure 3.6: We give 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Reg}}(\theta))$. The majority of the points lie below the line of slope one. The B-S model's errors are larger and more dispersed than the MT model's errors. See Section 3.5.3 for more details.

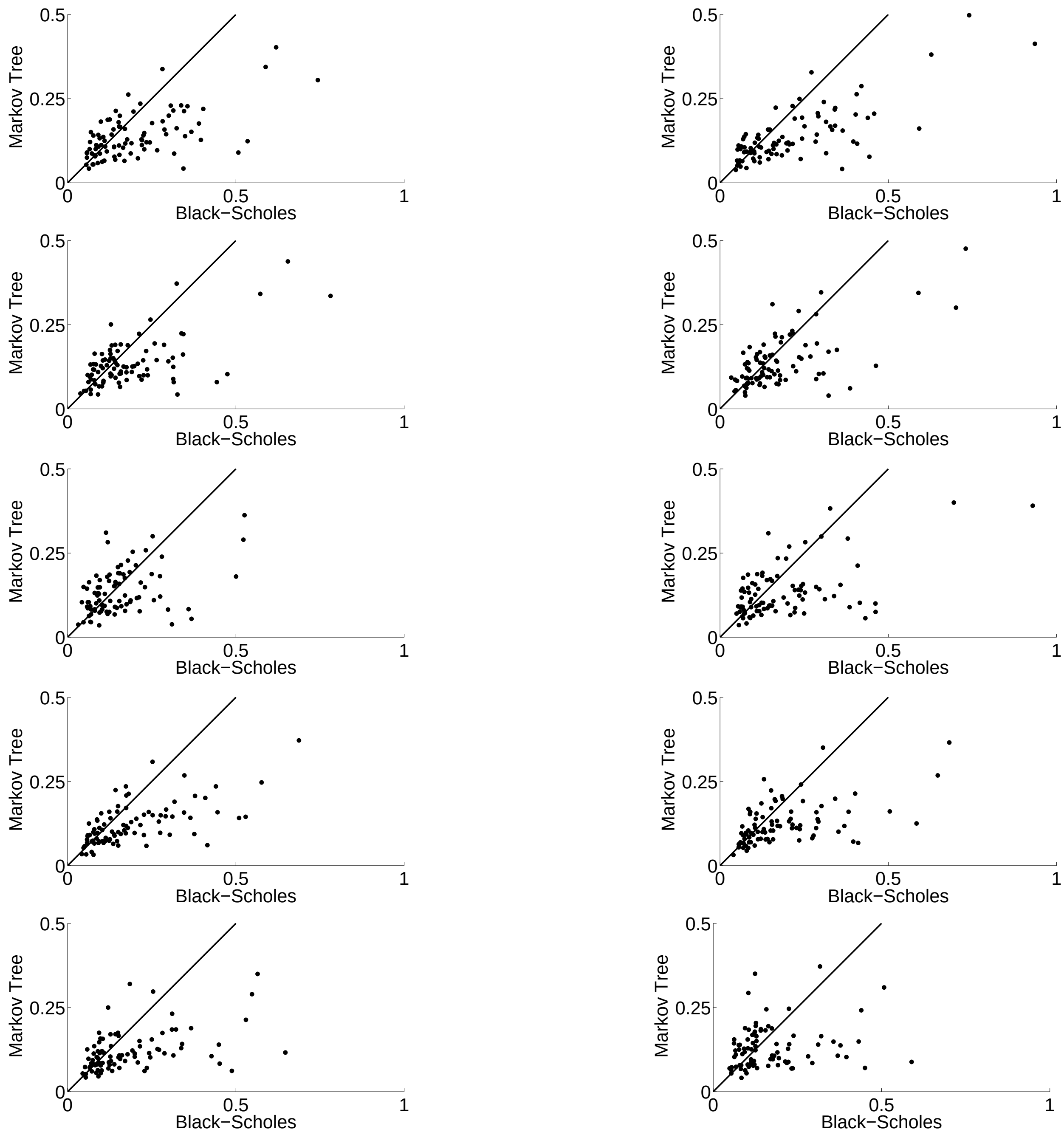

Figure 3.7: We give 10 scatterplots, one for each day of testing. Each scatterplot has 89 points of the form $(E^{\text{B-S}}(\theta), E^{\text{MT Naive}}(\theta))$. The majority of the points lie below the line of slope one. The B-S model's errors are larger and more dispersed than the MT model's errors. See Section 3.5.3 for more details.

Chapter 4

Large-Scale Empirical Testing

4.1 Introduction

Despite the prominence of option pricing models in the field of mathematical finance, most such models have never been subjected to empirical tests. In the academic literature, when a model's predictions are tested against data, it is typical to test only the out-of-sample pricing error using data consisting entirely of option contracts written on either the S&P 500 or S&P 100 indices (Bakshi et al., 1997; Nandi, 1996; Rubinstein, 1985; Bates, 2000, 1995; Corrado and Su, 1996; AitSahlia et al., 2010; Zhao and Hodges, 2012). In this Chapter, we focus on the question of which option pricing model achieves the best single instrument hedges, for options written on indices as well as individual equities. To answer this question, we use a large database of both individual equity options and index options to study the out-of-sample hedging performance of the Markov Tree (MT) model relative to two popular competing models, and we also substantially improve the statistical framework for fitting the MT model to observed data.

Recent work on both theoretical and empirical properties of the Markov Tree (MT) model have indicated that this model might perform well in large out-of-sample tests (Bhat and Kumar, 2010, 2012). While the model was originally proposed to explicitly account for the short-term dependence in an underlying asset's log returns (Bhat and Kumar, 2010), later work established a link between the MT model and option pricing models based on mixtures of normal distributions, justifying the application of the MT model to all individual equity options, not just those with first-order dependence in the log returns of the underlying stock (Bhat and Kumar, 2012). Empirical tests of the MT model against the classic Black-Scholes model (Black and Scholes, 1973) have been favorable thus far. The first test considered short- and mid-term European call options on six different stocks that were components of the CAC-40 index. Across 44 days of testing, the MT model outperformed the Black-Scholes model in out-of-sample pricing (Bhat and Kumar, 2010). The second test considered American call options on 89 stocks that were components of the S&P 100 index. In 10 days of testing, the MT model outperformed the Black-Scholes model in aggregate out-of-sample pricing error (Bhat and Kumar, 2012).

Previous work on the MT model did not explore the different methods for fitting the model to data, instead relying on *ad hoc* procedures based on historical volatilities (Bhat and Kumar, 2010, 2012). While these studies did consider genuine out-of-sample tests of pricing errors, the issue of hedging errors made by the MT model was left unadressed. Prior work on the MT model compared its predictions only to that of the Black-Scholes model, leaving out comparisons to more sophisticated models such as stochastic volatility models. Finally, while the data sets used in previous empirical tests were not small, they were not as large as data sets used in, e.g., the implied volatility literature.

In this Chapter, we make use of two databases of historical option prices. The first consists of 14,367 S&P 500 index call options from Jan. 1, 2009 to Dec. 31, 2010. The second consists of 3,599,468 unique LIFFE Paris equity call options traded between 19th September, 2009 and 18th June, 2012. Using this data, we compare the in-sample pricing errors, out-of-sample pricing errors, and out-of-sample hedging errors made by the Black-Scholes model (Black and Scholes, 1973), Heston's stochastic volatility model (Heston, 1993), and the Markov Tree model (Bhat and Kumar, 2010, 2012).

In order to carry out these tests, we develop three new methods for fitting the MT model to data, a problem that we frame as a nonlinear regression problem. The first two regression methods we develop are, respectively, underconstrained and overconstrained least-squares methods. The third method is a robust regression method that uses a pseudo-Huber loss function.

4.1.1 Results

Our primary result is that using any of the three regression methods developed in this Chapter, the MT model yields better out-of-sample hedging performance than either the Black-Scholes or Heston models.

For the overconstrained least-squares method, we develop a probabilistic simulation procedure to quantify the likelihood that the MT model will outperform Heston's model in repeated future trials. For the same overconstrained method, we analyze the regression residuals and show that they fit a generalized hyperbolic distribution with heavier-than-Gaussian tails, partly explaining why the robust regression method for fitting the MT model yields better results than the least-squares method.

There are a number of other insights that we take away from our analysis of the data. Regarding the general methods by which option pricing models are tested, we find that neither in-sample nor out-of-sample pricing errors by themselves are indicative of the out-of-sample hedging errors committed by any of the three models: Black-Scholes, Heston, or MT. In a similar vein, the results obtained by analyzing S&P 500 index options do not by themselves indicate what will happen when we analyze LIFFE individual equity options. Overall, while the volume of data we analyzed requires a nontrivial amount of computational time to process, the previous two observations indicate that our efforts yielded different conclusions than that of a more typical out-of-sample pricing test on index option data.

Other points we learned from this study concern the different methods to fit the MT model. The results obtained with the underconstrained least-squares method yield the smallest out-of-sample hedging errors. However, with this method, the model parameters are functions of the option's strike price and time to expiration, in conflict with the assumptions that go into the probabilistic derivation of the MT model. The overconstrained fitting methods (both least-squares and psuedo-Huber) do not assume that the model parameters depend on the strike price and time to expiration. These methods sacrifice a small improvement in hedging error for parsimony and interpretability.

4.1.2 Prior Work

The literature on option pricing is vast and has been surveyed elsewhere (Bates, 2003; Garcia et al., 2003; Broadie and Detemple, 2004). Here we focus our attention on work that empirically tests the hedging performance of one or more of the models studied in this Chapter.

To our knowledge, a comparison between the hedging performance of the Black-Scholes model, Heston's model, and any normal mixture distribution (NMD) model has only been carried out once before (Alexander et al., 2009). The NMD model that was tested, the Brigo-Mercurio model, is similar to the MT model in that both utilize log return distributions that are mixtures of normal distributions. Both models adopt a risk-neutral framework to derive closed-form option pricing formulas (Brigo and Mercurio, 2002). However, the Brigo-Mercurio model differs from the MT model in three key aspects:

1. The variances of the mixture components in the Brigo-Mercurio model have no interaction with one another; each variance is a function of a distinct model parameter. In the MT model, the variances of both mixture components interact strongly with one another; each variance is a function of the same two model parameters.

2. The Brigo-Mercurio model allows for an arbitrary number of mixture components, while the MT model allows for only two. When we restrict the Brigo-Mercurio model to two mixture components, the option price is a function of four model parameters (Alexander et al., 2009) rather than three for the MT model (Bhat and Kumar, 2010).

3. The procedure used to fit the Brigo-Mercurio model to data is different from the procedures described here or in past work on the MT model (Alexander et al., 2009; Bhat and Kumar, 2010, 2012).

These differences may serve to explain why the special case of the Brigo-Mercurio model tested in earlier work showed poorer hedging performance than either the Black-Scholes model or Heston's stochastic volatility model (Alexander et al., 2009). This contrasts sharply with the results presented in this Chapter, which show that the MT model's hedging performance is superior to that of the other two models.

Other empirical tests of Brigo-Mercurio NMD models have been carried out (Brigo et al., 2003; Alexander, 2004). The main focus of such works was to assess the in-sample fit of the NMD model's option prices using either different distributional assumptions on the components of the mixture (Brigo et al., 2003), or parameterizations of the NMD model that capture long- or short-term smile effects (Alexander, 2004). These tests have not addressed the issue of hedging performance, and have used small data sets consisting of option prices on one particular day.

In the finance literature, several authors have compared the hedging performance of stochastic volatility, jump diffusion, and Black-Scholes models (Nandi, 1996; Bakshi et al., 1997; Nandi, 1998; Bakshi et al., 2000; An and Suo, 2009; Kaeck, 2012). While this literature does include comparisons between varieties of stochastic volatility models, such as models featuring both jump diffusion and stochastic volatility, it is important to note that none of these papers include a multifactor non-stochastic volatility model in the suite of models being tested. This leaves open the question, addressed in the present work, of whether a constant volatility model that is more complex than the Black-Scholes model might outperform stochastic volatility models in out-of-sample hedging comparisons.

The empirical literature on stochastic volatility models typically relies on market data for options written on the S&P 500 index. One study, focused on hedging exotic options, studies options written on the EUR/USD Currency Option Volatility Index (An and Suo, 2009). To our knowledge, the present study is one of the first to use individual equity option data to study the hedging performance of Heston's stochastic volatility model.

The literature on empirical option pricing does include work that utilizes large databases of individual equity options. One of the earliest such works (Rubinstein, 1985) applies nonparametric tests to Chicago Board of Exchange (CBOE) individual equity option data to check for systematic differences between Black-Scholes prices and market prices. A later study examines a nearly two-year span of CBOE option data on 10 stocks (Lamoureux and Lastrapes, 1993) to test implications of the Hull-White stochastic volatility model (Hull and White, 1987).

A more recent study analyzed market data for options written on the S&P 100 index and the stocks that form its 30 largest components (Bakshi et al., 2003). Using 350,000 distinct option quotes (both calls and puts), this study examines the differences between implied risk neutral distributions for index and individual equity options. The effect of variables such as the price-to-earnings ratio and market capitalization on the skewness of implied risk neutral distributions has been studied using four years of end-of-week option data for 856 unique firms (Friesen et al., 2012)—this data set comprises 67,910 distinct option quotes.

Various studies have used individual equity option data to analyze various models for the implied volatility smile (Chou et al., 2011; Yan, 2011; Chang et al., 2012)—these studies each use between 14,120 and 400,000 distinct option quotes. In this Chapter, we analyze a data set that is an order of magnitude larger than the largest of the data sets we have seen mentioned in the literature.

In all the studies we have reviewed that use individual equity option data, none compute the hedging errors made by option pricing models. At the same time, the use of large data sets spanning years enables one to uncover long-term trends regarding model performance, trends that would have been missed by those using smaller data sets covering shorter periods of time.

4.2 Option Pricing Models

For a given option, let K denote the strike price, T the time in years to expiration, r the risk-free rate of interest, and S_0 the spot price of the underlying asset. We denote the option price as a function of the form $F(\mathbf{x}, \boldsymbol{\beta})$, where $\mathbf{x} = [K, T, r, S_0]$ and $\boldsymbol{\beta}$ is a vector of model parameters that must be statistically estimated from data. Typically, these model parameters include one or more volatilities.

We define moneyness as $m = S_0/K$, and maturity as the time to expiration in days.

Let us now review three classes of models. All models that we discuss treat r as constant over the life of the option.

4.2.1 Black-Scholes

The underlying asset price S_t at any time t is assumed to follow the SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{4.1}$$

where μ and σ are constants and W_t is the standard Wiener process. The annualized volatility σ is assumed to be constant. The Black-Scholes European call option price is

$$F^{\mathrm{BS}}(\mathbf{x}, \beta^{\mathrm{BS}}) = \Phi(d_1)S_0 - \Phi(d_2)K\exp(-rT), \tag{4.2}$$

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \qquad d_2 = d_1 - \sigma\sqrt{T}, \tag{4.3}$$

where Φ is the standard normal cumulative distribution function.

The Black-Scholes model is a one-parameter model with $\beta^{\mathrm{BS}} = \sigma$.

4.2.2 Heston

In Heston's stochastic volatility (SV) model, the underlying asset S_t is governed by the coupled system

$$dS_t = rS_tdt + \sqrt{v_t}S_tdW_t^s, \tag{4.4}$$

$$dv_t = \kappa(\theta - v_t)dt + \xi\sqrt{v_t}dW_t^v, \tag{4.5}$$

where v_t is the instantaneous variance of the asset price, dW_t^s and dW_t^v are Wiener processes with correlation ρdt, θ is the long variance, κ is the rate at which v_t reverts to θ, and ξ is the volatility of the volatility. Heston's model is a five-parameter model with $\boldsymbol{\beta}^{\mathrm{SV}} = (v_0, \rho, \theta, \kappa, \xi)$. The closed form European call option price for Heston's model is (Gilli and Schumann, 2010):

$$F^{\mathrm{SV}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{SV}}) = S_0P_1 - Ke^{-rT}P_2, \tag{4.6}$$

where

$$P_1 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\omega\log K}\phi(\omega - i)}{i\omega\phi(-i)}\right)d\omega, \qquad P_2 = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\omega\log K}\phi(\omega)}{i\omega}\right)d\omega \tag{4.7}$$

$$\phi(\omega) = e^Ae^Be^C \tag{4.8}$$

and

$$\begin{aligned}
A &= i\omega(\log S_0 + rT) \\
B &= \frac{\theta\kappa}{\xi^2}\left((\kappa - \rho\xi i\omega - d)T - 2\log\left(\frac{1 - g_2e^{-dT}}{1 - g_2}\right)\right) \\
C &= \frac{\frac{v_0}{\xi^2}(\kappa - \rho\xi i\omega - d)(1 - e^{-dT})}{1 - g_2e^{-dT}}, \quad \text{and} \\
d &= \sqrt{(\rho\xi i\omega - \kappa)^2 + \xi^2(i\omega + \omega^2)} \\
g_2 &= \frac{\kappa - \rho\xi i\omega - d}{\kappa - \rho\xi i\omega + d}.
\end{aligned}$$

4.2.3 Markov Tree

For $k \in \{-1, 0, 1\}$, let Z_k be a discrete random variable that achieves the outcomes $\{l_k, -l_k\}$ with probabilities $\{q_k, 1 - q_k\}$. Then, as originally proposed (Bhat and Kumar, 2010), the MT model assumes the underlying asset price S_n follows the persistent (or delayed) random walk

$$n = 1 : \log S_1 = \log S_0 + Z_0 \tag{4.9a}$$

$$n \geq 1 : \log S_{n+1} = \log S_n + \begin{cases} Z_1 & S_n \geq S_{n-1} \\ Z_{-1} & S_n < S_{n-1} \end{cases}. \tag{4.9b}$$

This process generates a risk-neutral probability mass function (pmf); using asymptotic analysis in an appropriate continuous-time limit, this pmf can be approximated very well by a mixture of normal densities, yielding the following expression for the price of a European call option (Bhat and Kumar, 2012):

$$\begin{aligned} F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})e^{rT} = q_0 S_0 \exp\left(\frac{\sigma_1^2}{2} + \mu_1\right)\Phi(x_1) + (1 - q_0)S_0 \exp\left(\frac{\sigma_2^2}{2} + \mu_2\right)\Phi(x_2) \\ - q_0 K\Phi(x_3) - (1 - q_0)K\Phi(x_4), \end{aligned} \tag{4.10}$$

where

$$x_i = \frac{\mu_i + \sigma_i^2 + \log(S_0/K)}{\sigma_i}, \quad x_{i+2} = \frac{\mu_i + \log(S_0/K)}{\sigma_i} \tag{4.11}$$

for $i \in \{1, 2\}$. In this Chapter, we treat (4.10) as the MT model's option price.

The MT model is a three-parameter model with $\boldsymbol{\beta}^{\mathrm{MT}} = (\sigma, \sigma_+, \sigma_-)$. The parameters q_k, l_k, σ_1, and σ_2 that appear either in the stochastic process (4.9) or on the right-hand sides of (4.10) and (4.11) are all functions of the components of $\boldsymbol{\beta}^{\mathrm{MT}}$—see Appendix B.1 for detailed algebraic expressions of these quantities.

4.3 Regression

We now describe how we use option price market data to fit the three models described in Section 4.2. Suppose that on day i we are interested in options on an underlying stock with symbol Θ.

Let $V_{\Theta,i}$ denote the column vector of all option prices associated with the underlying Θ on day i. Each row of $V_{\Theta,i}$ corresponds to a particular option contract, and each such contract corresponds to a row vector of the form $\mathbf{x}_j^{\Theta,i} = (K, T, r, S_0)$. Let $X^{\Theta,i}$ denote the matrix obtained by stacking these row vectors vertically, i.e.,

$$X^{\Theta,i} = \begin{bmatrix} \text{—} \mathbf{x}_1^{\Theta,i} \text{—} \\ \text{—} \mathbf{x}_2^{\Theta,i} \text{—} \\ \vdots \\ \text{—} \mathbf{x}_\nu^{\Theta,i} \text{—} \end{bmatrix},$$

where $\nu = |V_{\Theta,i}|$, the length of $V_{\Theta,i}$. Let $\mathbf{1} \in \mathbb{R}^\nu$ denote the column vector $\mathbf{1} = (1, 1, \ldots, 1)^\dagger$, where $\dagger$ denotes transpose. Then, once we fix the symbol Θ and the day i, the third and fourth columns of $X^{\Theta,i}$ are, respectively, $r\mathbf{1}$ and $S_0\mathbf{1}$; this is because the spot price of the underlying asset depends only on Θ and i, while the risk-free rate of interest depends only on i.

For the remainder of this section, we omit the Θ and i superscripts on X and $\mathbf{x}_j$—these superscripts will be used in Section 4.4.

For the data matrix X, and for each option pricing model $F(\mathbf{x}, \boldsymbol{\beta})$, we let denote $F(X, \boldsymbol{\beta})$ denote the result of applying $F(\mathbf{x}, \boldsymbol{\beta})$ to each row of X:

$$F(X, \boldsymbol{\beta}) = \begin{bmatrix} F(\mathbf{x}_1, \boldsymbol{\beta}) \\ F(\mathbf{x}_2, \boldsymbol{\beta}) \\ \vdots \\ F(\mathbf{x}_\nu, \boldsymbol{\beta}) \end{bmatrix}.$$

We can then formulate the nonlinear parametric regression problem

$$V_{\Theta,i} = F(X, \boldsymbol{\beta}) + \epsilon, \tag{4.12}$$

where ϵ is a column vector of residuals. The least-squares solution of this regression problem is

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\epsilon^\dagger \epsilon. \tag{4.13}$$

We now explain how special cases of (4.12) can be used to fit each of the option pricing models presented in Section 4.2.

4.3.1 Black-Scholes

Empirical studies reveal that allowing the regression coefficient or volatility $\beta^{\mathrm{BS}} = \sigma$ to depend on strike and time to expiration does not improve the hedging performance of the Black-Scholes model (Bakshi et al., 1997). For this reason, we take $V_{\Theta,i}$ to be all available call option prices for symbol Θ and day i, with X as the corresponding data matrix. We then set

$$V_{\Theta,i} = F^{\mathrm{BS}}(X, \beta) + \epsilon \tag{4.14a}$$

$$\beta^{\mathrm{BS}} = \arg\min_{\beta \in [0.05, 0.95]} \frac{1}{2}\epsilon^\dagger \epsilon, \tag{4.14b}$$

leading to a volatility that is independent of strike and time to expiration, a commonly used approach in prior empirical studies (An and Suo, 2009; Bakshi et al., 1997). To actually carry out the solution, we use the R function `optimize`, which combines golden section search and interpolation (R Core Team, 2012).

4.3.2 Heston

Since $\boldsymbol{\beta}^{\mathrm{SV}}$ is made up of five parameters, by analogy with linear regression problems, one does not expect the solution (4.13) to be unique unless $V_{\Theta,i}$ contains at least 5 rows. Therefore, a commonly used technique is to take $V_{\Theta,i}$ to be all available call option prices for symbol Θ and day i in a particular data set, so that the number of rows ν is always more than 5. The net effect of this is to compute via (4.13) a set of five parameters that do not depend on the option strike K and time to expiration T, i.e.,

$$V_{\Theta,i} = F^{\mathrm{SV}}(X, \boldsymbol{\beta}) + \epsilon \tag{4.15a}$$

$$\boldsymbol{\beta}^{\mathrm{SV}} = \arg\min_{\boldsymbol{\beta}\in\mathcal{S}} \frac{1}{2}\epsilon^{\dagger}\epsilon. \tag{4.15b}$$

Here $\mathcal{S} = \{(v_0, \theta, \rho, \kappa, \xi) \in \mathbb{R}^5 \mid 0.05 < v_0 < 0.95,\ 0.05 < \theta < 0.95,\ -0.9 < \rho < 0.9,\ 1 < \kappa < 6,\ 0.01 < \xi < 1.11\}$.

The main caveat of applying this procedure lies in the way (4.15b) is computed. The analytical gradient of the objective function—specifically F^{SV}—is not known, and numerically computed gradients are computationally expensive and inaccurate. This is because the evaluation of F^{SV} requires the numerical computation of an oscillatory integral. For these reasons, derivative-free rather than gradient-based optimization techniques are used to solve for $\boldsymbol{\beta}$ (Gilli and Schumann, 2010; Mikhailov and Nögel, 2003).

Two popular derivative-free techniques that are used to solve (4.13) for Heston's model are the Nelder-Mead algorithm (Fiorentini et al., 2002; West, 2005) and differential evolution (Gilli and Schumann, 2010). Using artificially created option data from a set of known parameters, (Gilli and Schumann, 2010) shows that differential evolution outperforms other derivative-free optimization techniques. We choose the Nelder-Mead algorithm for two reasons. First, differential evolution requires a prohibitively large number of evaluations of the objective function to achieve reasonable accuracy for a large-scale empirical test. Second, our tests on a subsample of LIFFE option data reveal that using Nelder-Mead with 500 iterations results in better convergence than differential evolution. The specific implementation of the Nelder-Mead algorithm we use is provided by the R function `optim` (R Core Team, 2012).

4.3.3 Markov Tree

We now describe three methods to fit the MT model to data. All three methods are implemented using gradient-based optimization, leveraging the smoothness of the MT model's option price with respect to the parameters $\boldsymbol{\beta}^{\mathrm{MT}}$. For all three methods, we defer any discussion of implementation details to Section 4.3.3.

Overconstrained L^2

The first method we consider is analogous to the procedure described above for Heston's stochastic volatility model. Using the full set of market option quotes for symbol Θ and day i, we formulate the regression problem and least-squares solution as

$$V_{\Theta,i} = F^{\mathrm{MT}}(X, \boldsymbol{\beta}) + \epsilon \tag{4.16a}$$

$$\boldsymbol{\beta}^{\mathrm{MT}} = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{2}\epsilon^{\dagger}\epsilon, \tag{4.16b}$$

where $\mathcal{B} = \{(x^1, x^2, x^3) \in \mathbb{R}^3 \mid 0.05 \le x^j \le 0.95,\ j = 1, 2, 3\}$. This formulation yields MT volatilities that do not depend on option strikes and expiration dates.

Overconstrained pseudo-Huber

Consider the loss function

$$L(x) = \delta \left(\sqrt{1 + \left(\frac{x}{\delta}\right)^2} - 1 \right), \tag{4.17}$$

adapated from the pseudo-Huber loss function (Jama, 2011). With this loss function, we can formulate an alternative solution to the regression problem (4.16a), one in which we replace the squared error loss function with $L(x)$:

$$\boldsymbol{\beta}^{\mathrm{MT}} = \underset{\boldsymbol{\beta} \in \mathcal{B}}{\arg\min}\, L(\epsilon), \tag{4.18}$$

where ϵ is defined as in (4.16a). The loss function L contains a parameter δ; for our empirical tests, we set $\delta = 0.01$, in which case $L(x)$ is a smooth approximation to the absolute value function $|x|$. Note that $\lim_{\delta \to 0} L(x) = |x|$, pointwise in x. Because L is less sensitive to outliers than the squared loss, the solution $\boldsymbol{\beta}^{\mathrm{MT}}$ may be thought of as a robust solution to the regression problem (4.16a). The smoothness of L enables us to apply gradient-based optimization techniques to solve (4.18).

Underconstrained L^2

We consider a special case of (4.12) in which we use only one row of the data matrix X and the corresponding row of the vector $V_{\Theta,i}$:

$$[V_{\Theta,i}]_j = F^{\mathrm{MT}}(\mathbf{x}_j, \boldsymbol{\beta}_{(j)}) + \epsilon. \tag{4.19}$$

Since $\boldsymbol{\beta}_{(j)}^{\mathrm{MT}}$ is a vector of three parameters, this problem is underconstrained, i.e., the set $\mathcal{F}_j = \{\boldsymbol{\beta} \mid [V_{\Theta,i}]_j = F^{\mathrm{MT}}(\mathbf{x}_j, \boldsymbol{\beta})\}$ is infinite. For this reason, we treat the nonlinear equation as a constraint, and solve the problem

$$\boldsymbol{\beta}_{(j)}^{\mathrm{MT}} = \underset{\boldsymbol{\beta} \in \mathcal{F}_j}{\arg\min}\, \|\boldsymbol{\beta} - \mathbf{b}\|_2^2, \tag{4.20}$$

where $\mathbf{b} = (0.5, 0.5, 0.5)$. This formulation yields a strike- and expiration-dependent set of MT volatilities.

Recall from Section 4.2.3 that the parameter vector $\boldsymbol{\beta}^{\mathrm{MT}}$ can be used to calculate three risk-neutral probabilities $\{q_k\}_{k=-1}^{1}$. Let $\mathcal{Q} = \{\boldsymbol{\beta} \in \mathbb{R}^3 \mid 0 < q_k(\boldsymbol{\beta}) < 1,\ -1 \leq k \leq 1\}$. For all index options considered in this study, when we find the solution (4.20) and insert it into (4.19), the residual error ϵ is zero to machine precision. We also find that $\boldsymbol{\beta}_{(j)}^{\mathrm{MT}} \in \mathcal{Q}$, i.e., the q_k's are valid probabilities.

For approximately 10^3 individual equity options considered in this study, solving (4.20) yields $\boldsymbol{\beta}_{(j)}^{\mathrm{MT}} \notin \mathcal{Q}$. For only these options, we solve the following problem instead of (4.20):

$$\boldsymbol{\beta}_{(j)}^{\mathrm{MT}} = \underset{\boldsymbol{\beta} \in \mathcal{F}_j \cap \mathcal{Q}}{\arg\min}\, \|\boldsymbol{\beta} - \mathbf{b}\|_2^2. \tag{4.21}$$

In practice, this yields solutions that satisfy both the $\mathcal{F}_j$ and $\mathcal{Q}$ constraints.

Implementation Details

To obtain either of the overconstrained solutions (4.16b) or (4.18), we use the L-BFGS-B algorithm (Byrd et al., 1995). This is a quasi-Newton solver that uses a limited memory (L) version of the BFGS update formula, while also handling box constraints (B). Our use of a quasi-Newton solver means that we avoid calculating the exact Hessian of the objective function. The L-BFGS-B implementation we use is built into the `optim` command in R (R Core Team, 2012).

To obtain either of the underconstrained solutions (4.20) or (4.21), we use the package `nloptr` (Ypma, 2011), an R interface to the `nlopt` package (Johnson, 2013). Specifically, we use this package's implementation of an augmented Lagrangian method (Conn et al., 1991; Birgin and Martinez, 2008).

For all of the codes/algorithms just mentioned, we pass user-defined functions that use exact formulas to compute the gradient of the MT option price F^{MT} with respect to $\boldsymbol{\beta}^{\mathrm{MT}}$. This gradient is given in detail in Appendix B.2.

For the overconstrained methods, our results show that at a computed optimum $\boldsymbol{\beta}^{\mathrm{MT}}$, the gradient of the objective function is near zero. We also find that either the Hessian at the optimum is positive definite, or the computed optimum lies on the boundary of the feasible set $\mathcal{B}$.

For the underconstrained methods, we rely on the optimization algorithm to single out a unique element of the feasible set. That is, we are less interested in whether the algorithm converges to a local minimum, and more interested in how well the constraints are satisfied. In all of our tests, the solution of either (4.20) or (4.21) yields (i) valid risk-neutral probabilities $\{q_k\}_{k=-1}^{1}$ and (ii) residual errors in (4.19) that are zero to at least four decimal places, sufficient for the purposes of this study.

4.4 Tests

In the previous section, we described five procedures for fitting an option pricing model to market data: one Black-Scholes procedure, one stochastic volatility procedure, and three MT procedures. All five procedures can be viewed as special cases of (4.12), which, for an option on the underlying Θ that is being priced on day i, reads

$$V_{\Theta,i} = F(X^{\Theta,i}, \boldsymbol{\beta}^{\Theta,i}) + \epsilon_{\Theta,i}. \tag{4.22}$$

Here we have highlighted the fact that the data matrix X, the regression coefficients $\boldsymbol{\beta}$, and the residuals ϵ all depend on Θ and i. In what follows, we refer to $\epsilon_{\Theta,i}$ as the *in-sample pricing errors*.

4.4.1 Out-of-Sample Pricing Error

Having computed $\boldsymbol{\beta}^{\Theta,i}$, we can use this vector of regression coefficients to price options on the underlying Θ on day $i+1$. This leads to the one day *out-of-sample pricing error* vector

$$e_{\Theta,i+1} = V_{\Theta,i+1} - F(X^{\Theta,i+1}, \boldsymbol{\beta}^{\Theta,i}). \tag{4.23}$$

Note that $V_{\Theta,i+1}$ consists of all available call option prices in our data set for day $i+1$ and symbol Θ. The data matrix $X^{\Theta,i+1}$ is such that the risk-free interest rate and spot prices are current as of day $i+1$, while the strikes and the expiration dates are read from the call option contracts available on day $i+1$ for symbol Θ.

4.4.2 Out-of-Sample Hedging Error

Hedging is the process of creating a risk-free portfolio consisting of risky assets. Owing to transaction costs and other financial considerations, a simple and practical form of such a portfolio is one which uses the minimum number of financial instruments (Bakshi et al., 1997). Hedging with a portfolio consisting only of an option and shares of its underlying is commonly referred to as single instrument hedging.

Consider a portfolio created by selling *one* call option at the price $V_{\Theta,i}^{K,T}$ and buying n shares of it's underlying at the price of S_0 per share. The residual cash obtained through this transaction at time $t=0$ on day i is $\pi_{\Theta,i}^{K,T} = V_{\Theta,i}^{K,T} - nS_0$. The value of this portfolio depends on the market price of the call option and the market price of its underlying. A change in the price of the underlying at time $t=\Delta t$ leads to a change in the call option price and the value of the portfolio. We seek a portfolio whose value is insensitive to small changes in the underlying price. For the BS and MT models, this can be achieved by choosing n such that

$$\frac{\partial \pi_{\Theta,i}^{K,T}}{\partial S_0} = \frac{\partial V_{\Theta,i}^{K,T}}{\partial S_0} - n = 0.$$

Approximating the market price $V_{\Theta,i}^{K,T}$ by the model price $F(\mathbf{x}_j^{\Theta,i}, \boldsymbol{\beta}^{\Theta,i-1})$, we have

$$n = \frac{\partial V_{\Theta,i}^{K,T}}{\partial S_0} \approx \frac{\partial F(\mathbf{x}_j^{\Theta,i}, \boldsymbol{\beta}^{\Theta,i-1})}{\partial S_0}. \tag{4.24}$$

With this n, the portfolio described above with value $\pi_{\Theta,i}^{K,T}$ is often called a *delta neutral* portfolio. For the MT model, we report the exact forms of n in Appendix B.3. Since the only stochastic term in the Black-Scholes and MT models is the stock price, a delta neutral portfolio can be created using a single option and its underlying. This cannot be done for a stochastic volatility model, as the value of the portfolio is driven not just by the stock price but also the stochastic volatility. Nevertheless, for Heston's model, and for a portfolio consisting of a call option and shares of its underlying, we can choose n such that the variance of the portfolio is minimized (Bakshi et al., 1997). This is called a minimum variance hedge.

Equation (4.24) represents continuous rebalancing—the number of shares n have to be continuously changed. As this is not possible, in practice, a portfolio is rebalanced (equivalently, n is adjusted) only at discrete times, typically each day at the end of trading.

To evaluate the hedging performance of different models, we first construct a risk-free portfolio at time $t=0$ (day i) consisting of an option (with strike K and time to expiration T) and n shares of its underlying (with symbol Θ). To calculate this n, we use (4.24) with $\boldsymbol{\beta}$ computed using data on day $i-1$. We invest the residual cash generated into a risk-free bond maturing at time $t=\Delta t$ on day $i+1$. We assess the value of the stock and option at time $t=\Delta t$ (day $i+1$) by closing the portfolio, thereby generating a cash value of $nS_{\Delta t} - V_{\Theta,i+1}^{K,T-\Delta t}$. Maturity of the bond generates another $\pi_{\Theta,i}^{K,T} e^{r_i \Delta t}$ on day $i+1$. The hedging error of this self financed portfolio created on day i and liquidated on day $i+1$ is then

$$H_{\Theta,i}^{K,T} = \pi_{\Theta,i}^{K,T} e^{r_i \Delta t} - V_{\Theta,i}^{K,T-\Delta t} + n_{\text{model}} S_{\Delta t}. \tag{4.25}$$

This procedure of using data on day $i-1$ to compute $\boldsymbol{\beta}$ and n, form the portfolio on day i, and liquidate the portfolio on day $i+1$ follows prior work (Bakshi et al., 1997).

4.5 Data

4.5.1 LIFFE Paris individual equity options

Market data on different LIFFE contracts traded electronically is available for download from the website `http://www.liffe.com/data/`. Using a Python script, we download data everyday for all contracts traded on the LIFFE exchange and stack it in a MySQL database. In this way, we build sixteen different databases based on different markets and different contract types. To keep the analysis feasible, we consider only the Paris Equity Options data for this study. We consider all options contracts written on all stocks on the LIFFE Paris Equity Options traded between September 18, 2009 and June 18, 2012, encompassing 707 trading days worth of data. To further reduce the size of the data set and keep computational time tractable, we only consider call options traded within this period, leaving 7,361,451 unique options. We then apply standard filtering techniques to improve the fitting process for different models, and to remove the bias involved in pricing options that are not traded. Specifically, we remove from our data set

- short-term options, i.e., options with maturity strictly less than seven trading days,
- deep in-the-money options (with moneyness $S_0/K \geq 1.4$), and
- deep out-of-the-money options (with moneyness $S_0/K \leq 0.8$),

leaving us with 3,483,461 call options in the LIFFE data set. This number of unique options is approximately 100 times greater than the number of options analyzed in prior empirical hedging studies (Bakshi et al., 1997; Nandi, 1996; Zhao and Hodges, 2012).

For this data set, the option price that we use is the LIFFE settlement price. Settlement prices are determined using the trade-weighted average market value of the option together with a variety of technical considerations spelled out in LIFFE guidelines (LIFFE, 2006). Using these settlement prices avoids pitfalls associated with daily closing prices that have been documented in the literature (Rubinstein, 1985).

In Table 4.1, we report the number of options on (i) all 118 LIFFE option symbols and (ii) 25 LIFFE option symbols with non-dividend-paying underlying in our database categorized by moneyness-maturity. In Table 4.2, we also report the average price of options in each of the above categories. We note that, for all moneyness-maturity categories, the average price of options on stocks that do not offer dividends is smaller than the average price of all options.

4.5.2 SPX options

CBOE market data on traditional European style options on the S&P 500 index for 2009 and 2010 is available from `http://www.deltaneutral.com`. Options on the S&P 500 index have been considered in empirical studies before (Bakshi et al., 1997, 2000; Nandi, 1996) and are known to be the benchmark options to test European option pricing models (Rubinstein, 1985; Nandi, 1996). Again, we follow standard filtering techniques (Kaeck, 2012); after removing all put options, we further remove

- short-term options, i.e. options with maturity strictly less than seven trading days,
- deep in-the-money options (with moneyness $S_0/K \geq 1.3$),
- deep out-of-the money options (with moneyness $S_0/K \leq 0.8$), and
- options with zero trading volume,

leaving us with an overall SPX data set made up of 5,683 call options for 2009, and 14,367 call options in 2010. This data set consists of bid and ask prices; following standard convention (Bakshi et al., 1997; Kaeck, 2012), we take the midpoint of the bid and ask prices to be the the market option price.

We report the number of SPX options categorized by moneyness and maturity in Table 4.3. Here, we note that after filtering, there are no long-term options (i.e., with maturity exceeding 180 days) in 2010. In Table 4.4, we present the average price of S&P 500 index options categorized by moneyness and maturity.

4.5.3 Interest rates

We use 90-day LIBOR rates as our proxy for the risk-free rate of interest. The risk-free rate for any option contract dated in any given month is assumed to be the 90-day LIBOR rate at the beginning of the corresponding month.

4.5.4 Dividends

During the period of study, 2009–2012 for LIFFE options and 2009–2010 for SPX options, dividends paid by either LIFFE or S&P 500 equities were either zero or negligible.

4.6 Results

In this section, we present results on four data sets:

- all LIFFE data as described in Section 4.5—see Tables 4.5 and 4.9.
- the subset of the LIFFE data consisting of options on 25 non-dividend-paying stocks—see Tables 4.6 and 4.10.
- SPX data from calendar year 2009—see Tables 4.7 and 4.11.
- SPX data from calendar year 2010—see Tables 4.8 and 4.12.

Our primary result is that, for any of these data sets, any of the regression procedures for fitting the MT model described in Section 4.3.3 yield smaller out-of-sample hedging errors than either Heston's model or the Black-Scholes model.

All entries of Tables 4.5-4.12 are computed in the following way. First, we compute the appropriate error (i.e., in-sample, out-of-sample pricing, out-of-sample hedging) for each option in the indicated data set. We then bin options into moneyness-maturity categories. We include "Overall" bins that denote either all options, options binned only by moneyness, or options binned only by maturity. Finally, we calculate the mean absolute error in each bin. All LIFFE errors are in Euros (€), while all SPX errors are in US dollars ($).

Note that an option must exist in our data set on both day i and day $i + 1$ in order for a hedging error to be calculated.

4.6.1 Comparison of different option pricing models

In our first set of results, we compare the Black-Scholes model, Heston's stochastic volatility model, and the MT model using the overconstrained L^2 fitting procedure from Section 4.3.3.

In-Sample Pricing Errors. The first panel of Table 4.5 shows the in-sample errors for all three models on the entire LIFFE data set. Heston's model features an overall error (€0.1981) that is €0.02 less than the overall error committed by the second best model, the MT model. This is not surprising; in the framework of nonlinear regression, one expects a model with two extra parameters to provide a better in-sample fit.

Note that the MT model outperforms Heston's model for long-term options, i.e., those with maturity greater than 180 days. Two possible causes for this are (i) the necessity of using derivative-free optimization to fit Heston's model, and (ii) the possibility that the MT model fits option data on dividend-paying stocks better than Heston's model.

We contrast the results on the entire LIFFE data set with results on the subset of the LIFFE data consisting of options on stocks that pay no dividend. From the first panel of Table 4.6, we see that Heston's model performs much better than both the Black-Scholes and MT models, beating the MT model in overall performance by about €0.05. Comparing the in-sample portions of Tables 4.5 and 4.6, we see that the three models reduce their overall errors by 44.37% (Black-Scholes), 46.16% (MT), and 66.18% (Heston). Based on this, we hypothesize that Heston's model performs better when there is complete information about the market, e.g., when dividend data is available.

Turning to Tables 4.7 and 4.8, we see one of the reasons for working with a large database of individual equity options: the results for index options can differ. Looking at the overall in-sample errors in these tables, we see that Heston's model outperforms the MT model for 2009, but that the MT model outperforms Heston's model for 2010. Unlike the results for LIFFE options, the SPX option results show that Heston's model fit long-term options particularly well for both 2009 and 2010 index options. This last result confirms tests carried out in the litearture (Bakshi et al., 1997).

Out-of-Sample Pricing Errors. The second panels of Tables 4.5-4.8 show out-of-sample pricing errors for each of the four data sets described earlier.

For LIFFE equity options, whether restricted to non-dividend-paying underlying or not, Heston's model yields smaller overall errors than either the Black-Scholes or MT models, and yet for

SPX index options for either 2009 or 2010, the MT model's overall errors are the smallest. We again see a difference in model performance between index and individual equity options.

It is clear from the results that all models' out-of-sample pricing errors exceed their in-sample pricing errors; for the full LIFFE data set, specifically, the overall out-of-sample errors are larger by 1.4%, 2.3%, and 4.8% than the overall in-sample errors for the Black-Scholes model, the MT model, and Heston's model, respectively. For the 2009 SPX data set, the model-wise increases in overall pricing errors from in-sample to out-of-sample tests are 23.53%, 54.11%, and 139.8%, while for the 2010 SPX data, these model-wise increases are 22.99%, 23.27%, and 27.94%.

There are two trends that we note here. First, the out-of-sample results for all models are much closer to the in-sample results for LIFFE individual equity options than they are for S& P 500 index options. This shows, again, that it is useful to test option pricing models on both types of data sets.

Second, for both LIFFE and SPX data sets, Heston's model consistently has the largest percentage increase in overall pricing errors from in-sample to out-of-sample tests. This leads us to hypothesize that Heston's model may be overfitting; we find much stronger support for this in our hedging results below.

Since LIFFE and SPX results do not show consistency in the reduction of out-of-sample pricing errors as a function of model complexity (i.e., number of parameters), we turn to another out-of-sample test to determine if Heston's model either overfits the data or is truly superior, as the in-sample results indicate.

Out-of-Sample Hedging Errors. Out-of-sample hedging errors are displayed in the final panels of Tables 4.5-4.8, for each of the four data sets described earlier.

On the entire LIFFE data set, the MT model's overall hedging errors are 41.88% lower than that of Heston's model, and 301.7% lower than that of the Black-Scholes model. For the subset of LIFFE data consisting of options on non-dividend-paying underlying, the MT model's overall hedging errors are 41.32% and 296.84% lower than for Heston's model and the Black-Scholes model, respectively. For SPX options from 2009 (2010), the MT model's overall hedging errors are 57.68% (69.47%) and 86.5% (71.13%) smaller than what we find for Heston's model and the Black-Scholes model, respectively.

There are several insights that we obtain from these results. First, and most importantly, neither the in-sample nor the out-of-sample pricing performance is indicative of the out-of-sample hedging performance of an option pricing model. Second, among all the performance metrics considered, the hedging errors seem to be least affected by not accounting for the dividends in the option pricing models—this is indicated by the consistency of the hedging results across Tables 4.5 and 4.6.

We note another consistent trend in the out-of-sample hedging results in all four Tables 4.5-4.8. For the MT model, if we bin errors *only* by maturity, then short-term options (i.e., with maturity less than 60 days) yield the best hedging performance; similarly, if we bin errors *only* by moneyness, then the most out of the money options (i.e., with $m < 0.94$) yield the best hedging performance.

Next, we visualize and assess hedging errors in a different way. Suppose our data set is one of the four data sets described at the beginning of this section. For a given stock symbol Θ and a given day i, we sum the *raw* hedging errors due to all different options available on day i on the underlying Θ. This yields a hedging error for symbol Θ on day i. We sum over Θ, and thereby obtain a time series of *market hedging errors*. As there are four data sets, we obtain four time series for each of the three models that were tested.

In Figure 4.1, we plot these four time series for Heston's model (blue) and the MT model (red). The time series for the Black-Scholes model is omitted, because it increases the vertical scale of the plot to an extent that we miss details in the blue and red curves. The figure indicates that the red curves are enveloped by the blue cruves, meaning that on each day, it is usually the case that the market hedging error is larger for Heston's model than for the MT model. For the full LIFFE data set (respectively, the non-dividend-paying LIFFE data set), the MT market hedging error is smaller in absolute value than that of Heston's model for 603 (579) out of 705 days. For the SPX data set, and for both 2009 and 2010, the MT market hedging error is smaller in absolute value than Heston's model's market hedging error 269 days out of a total of 354 days. Again we see remarkable consistency across all four time series; in all four cases, the empirical probability that the MT model yields smaller hedging errors is between 0.76 and 0.85.

Apart from showing that the MT model outperforms Heston's, these plots also reveal how the model's daily hedging errors vary on a daily basis. This variation has not been plotted before, even in large-scale empirical studies. The plots clearly show that there are several days, e.g., near 09-2011 for the LIFFE plots and 01-2009 and 05-2010 for the SPX plots, where the market hedging error for Heston's model is much larger than for the MT model. On the other hand, when Heston's model has a smaller hedging error than the MT model, the difference is small. These fine-grained results are important from a risk management perspective, especially if one seeks a model that never "blows up." An overly narrow focus on errors averaged across time obfuscates this point.

We form two overall conclusions from the out-of-sample hedging results. First, because the MT model consistently produces the least out-of-sample hedging errors, both in the overall categories and in almost all moneyness-maturity bins, the MT model should be used for risk management purposes rather than the other two models studied. Second, the MT model achieves its superior hedging performance with two fewer regression coefficients than Heston's model. While the in-sample pricing errors decrease as a function of model complexity, the out-of-sample hedging errors are minimized by a model with three parameters (MT) rather than five (Heston), leading us to believe that Heston's model does indeed overfit the data.

4.6.2 Performance of MT model regression procedures

Next, we compare the three methods for fitting the MT model described in Section 4.3.3. The layout of results in Tables 4.9-4.12 follows that of Tables 4.5-4.8; the main difference is that the three models considered previously are replaced by three methods for fitting the MT model: the overconstrained L^2 method, the overconstrained pseudo-Huber method, and the underconstrained L^2 method.

In-Sample Pricing Errors. In-sample pricing errors for each of the four data sets described above are shown in the first panels of Tables 4.9-4.12. The underconstrained L^2 method has residual errors that are zero to four decimal places; these values are omitted here. Focusing our attention on the two overconstrained methods, we see that in overall error for the four data sets considered, the pseudo-Huber method is better than the L^2 method by €0.0094, €0.0067, €0.0336, and

€0.049. However, examining each individual moneyness-maturity bin, we see that the L^2 method has a smaller error for many bins, indicating that neither method is clearly superior in terms of in-sample fit.

Out-of-Sample Pricing Errors. In the second panels of Tables 4.9-4.12, we present out-of-sample pricing errors for the three regression methods and each of the four data sets. The underconstrained L^2 scheme clearly outperforms all other regression procedures across all four data sets. This is true not only for overall errors, but is also true for nearly every moneyness-maturity category. Moreover, comparing the performance of the underconstrained L^2 MT method to that of Heston's model on each of the four data sets, we see that the MT method consistently produces significantly smaller out-of-sample pricing errors.

These results alone justify our inclusion of the underconstrained L^2 method. While strike- and maturity-dependent volatilities are inconsistent with the assumptions in the MT stochastic model, they do lead to the smallest out-of-sample errors. From a practitioner's point of view, we expect the underconstrained L^2 method to be the method of choice for fitting and using the MT model.

Between the two overconstrained methods, the pseudo-Huber method has smaller overall errors on the LIFFE data sets and on the 2010 SPX data set, while the L^2 method has smaller overall errors on the 2009 SPX data set. For LIFFE options, if we focus our attention on short- and medium-term options, the pseudo-Huber method performs noticeably better than the L^2 method; the L^2 method is the superior method for long-term options. These statements do not carry over to the SPX data set, indicating again the difference between tests for individual equity options and index options.

Out-of-Sample Hedging Errors. The third panels of Tables 4.9-4.12 show the out-of-sample hedging errors for each of the four data sets and each of the three regression methods. As with the out-of-sample pricing errors, the underconstrained L^2 method yields smaller overall errors across all four data sets.

The overall hedging errors for the overconstrained L^2 and the overconstrained pseudo-Huber methods are greater than the overall hedging errors for underconstrained L^2 method by 10.65% and 9.69%, respectively, for 2009 SPX options and 29.34% and 27.38%, respectively, for 2010 SPX options. For both LIFFE data sets, the underconstrained L^2 method yields overall errors that are between 15.6% and 16.5% smaller than with either of the overconstrained methods.

These results show that the underconstrained L^2 regression MT method produces the least out-of-sample hedging errors across all data sets, not just overall, but also in each individual moneyness-maturity bin.

Following this procedure, a practitioner can choose a unique option with which to form a risk-free portfolio, and then estimate the MT model parameters for this particular option using the method outlined in section 4.3.3. This circumvents the need to collect data for all options traded on a given day for a given symbol, a requirement for the overconstrained methods.

From these results, we also note that the overconstrained pseudo-Huber method slightly outperforms the overconstrained L^2 method, in the overall category, for all four data sets.

The results help us draw two final conclusions. First, from a practical perspective, the underconstrained L^2 regression MT model provides the least out-of-sample hedging errors. This procedure allows the MT model parameters to depend on the strike and maturity of the option, leading to an increase in the number of model parameters. While we have conducted large scale out-of-sample empirical tests to guard against drawing conclusions from the in-sample fit of this method, we note that this procedure does not conform with MT model assumptions that do not allow the stock price process, and, in turn, the model parameters, to depend on option's strike and maturity. Second, the overconstrained pseudo-Huber method, an example of a robust nonlinear regression procedure, produces the least out-of-sample hedging errors among all overconstrained regression procedures carried out in this Chapter. This regression procedure is consistent with MT model assumptions, and does not require more time to run than the standard overconstrained L^2 method. While the improvement over the overconstrained L^2 method may be slight, the results lead us to hypothesize that *some* robust regression technique for fitting the MT model may yield a much larger improvement over the L^2 method.

4.7 Error Analysis

In order to achieve a better quantitative understanding of the superior hedging results displayed by the MT model in Section 4.6, we analyze the MT model's errors. First, we show that for the overconstrained L^2 MT model, the tails of the in-sample residual distribution decay more slowly than the tails of the normal distribution, helping to explain why the robust pseudo-Huber regression procedure for fitting the MT model yields marginally better results. Second, we show using a statistical simulation procedure that the MT model's superior hedging performance is not likely to be due to chance, but instead due to the model's robustness with respect to noise in option data.

In this section, we restrict our attention to the SPX data set, enabling us to run a reasonable number of simulations. This is justified by the results from Section 4.6, which show consistency of MT model results across LIFFE and SPX options. In what follows, the underlying Θ will be the S&P 500 index, rather than an individual equity.

4.7.1 In-Sample Error Analysis

While fitting the regression model as in (4.22), we obtain the residual vector $\epsilon_{\Theta,i}$ for each day i. When we use the overconstrained L^2 method from Section 4.3.3, the regression is performed under the assumption that the residuals are (i) independent of option strike and maturity, and (ii) independent and identically distributed (i.i.d.) samples from an error random variable. We consider data from 2009 and 2010 separately since our data shows noticeably different liquidity of options for these two years that is also captured in our results.

Collecting all error vectors $\epsilon_{\Theta,i}$ for 2009 and 2010, we obtain, respectively, 5,647 and 14,366 i.i.d. samples of the random variables E_{2009} and E_{2010}. We then subject our samples of E_{2009} and E_{2010} to exploratory data analysis. We then fit a generalized hyperbolic distribution (GHD) to both these samples and find the maximum likelihood estimates via expectation maximization. We fit a GHD for three reasons. First, the GHD has been used with success in finance where heavier-than-normal tails arise (McNeil et al., 2005). Second, the GHD includes as special cases, other distributions of interest: hyperbolic, normal inverse Gaussian, variance gamma, student t, and normal that enables us to perform likelihood ratio tests. Third, the best fit GHD is a very good match for the kernel density estimates of E_{2009} and E_{2010}—see Fig 4.2).

We use the five parameter parametrization of the GHD (Luethi and Breymann, 2013, section 4.2). The parameters of the GHD that best fits E_{2009} and E_{2010} are given in Table 4.13.

We employ likelihood ratio tests and AIC model selection to test the error distribution. For the samples of E_{2009}, the likelihood ratio tests reject the hypothesis that the true underlying distribution belongs to four of the five special cases of the GHD mentioned above—the exception is the normal inverse Gaussian distribution, with a p-value of 0.05. For the samples of E_{2010}, the likelihood ratio test rejects that the true underlying distribution belongs to any of the five specical cases of the GHD mentioned above. For both E_{2009} and E_{2010} samples, AIC model selection criteria also selects GHD as the best fitted model for the 2010 errors.

Taken together, these exploratory results indicate that the GHD, rather than any of the five special cases we tested, is a good fit for the residuals from 2009 and 2010. Note that the tail decay in the GHD is given by $|x|^a e^{mx}$, where $a = \lambda - 1$ and m differs for the left and right tails. For our fitted distribution, a is -1.4 (0.22) for 2009 (2010) respectively. The exponent m is 0.56 (-0.472) for the left (right) tail for 2009 and 1.048 (-0.49) for the left (right) tail for 2010. This indicates that the fitted GHD is asymmetric and has heavier tails than the normal distribution.

The tails of the GHD fitted to the 2009 and 2010 residuals are inconsistent with the assumption that the error ϵ in (4.16a) is normally distributed. If ϵ does indeed have heavier-than-normal tails, then the least-squares solution (4.16b) will not be the maximum likelihood estimator of $\boldsymbol{\beta}$ in (4.16a). On the other hand, the pseudo-Huber solution (4.18) will be close to the minimizer of $\|\epsilon\|_1$, which is the maximum likelihood estimator of $\boldsymbol{\beta}$ when the error ϵ has a Laplace distribution with asymptotic tail decay $e^{-b|x|}$. The tail decay of the fitted GHD is closer to Laplace than normal, helping to explain why the pseudo-Huber method performs marginally better than the L^2 method.

4.7.2 MT Model Performance: Perturbed Regression Coefficients

Let j be a fixed day in either 2009 or 2010, and let $e_{\Theta,j}$ denote a vector of samples from the GHD using the best fit parameters from Table 4.13 for the appropriate year. We assume $e_{\Theta,j}$ has the same number of components as $V_{\Theta,j}$ in (4.22), i.e., the number of available call options for underlying Θ on day j. Let $\boldsymbol{\beta}^{\Theta,j}$ denote the MT regression coefficients computed using (4.16b) for underlying Θ on day j. Then define the vector $\mathcal{V}^{\Theta,j}$ of simulated option prices by

$$\mathcal{V}^{\Theta,j} = F^{\mathrm{MT}}(X^{\Theta,j}, \boldsymbol{\beta}^{\Theta,j}) + e_{\Theta,j}. \tag{4.26}$$

By comparison with (4.22), we see that if ϵ and e have the same distribution, then $\mathcal{V}$ and V have the same distribution as well.

We now use $\mathcal{V}^{\Theta,j}$ in place of the market prices $V^{\Theta,j}$ in (4.16a) and obtain a new set of regression coefficients $\widetilde{\boldsymbol{\beta}}^{\Theta,j}$ using (4.16b). Using the coefficients $\widetilde{\boldsymbol{\beta}}^{\Theta,j}$ on day $j = i-1$, we compute out-of-sample hedging errors using a self-financed portfolio created on day i and liquidated on day $i + 1$ as in Section 4.4.2. Repeating this procedure 50 times, and aggregating the data across each year, we obtain $282,350$ samples of H^{MT}_{2009} and $718,300$ samples of H^{MT}_{2010}, where H^{MT}_y is a random variable representing the out-of-sample hedging error made by the MT model on one day in year y. Note that we repeat the procedure 50 times so that we have enough samples of the random variable τ_y (explained next) such that the mean of τ_y converges to a constant value.

To relate the MT model's performance to that of Heston's model, we compute $\tau_y = |H^{\mathrm{SV}}_y| - |H^{\mathrm{MT}}_y|$, where H^{SV}_y is the out-of-sample hedging error for Heston's model computed as in Section

4.4. In Figure 4.3 we show kernel density estimates (KDE) of τ_y for $y = 2009$ and $y = 2010$. The significant asymmetry present in both years' KDE plots indicate that the MT model's superior hedging performance persists even when market prices differ from historical market prices. To quantify the improvement in performance, we report the deciles of τ_y in Table 4.14. From the table, it is clear that only about 20% of SPX options can be hedged better using Heston's model. The mean of τ_{2009} and τ_{2010} are \$0.8707 and \$0.9089 respectively. Aslo, the empirical cumulative distribution function value of τ_{2009} and τ_{2010} at zero is \$0.2366 and \$0.2141 respectively. The performance of the MT model indicates its robustness with respect to noisy option prices.

	All 118 stocks				25 non-dividend paying stocks			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	281602	280725	411781	974108	58022	55454	78524	192000
0.94–0.97	98580	65437	97078	261095	16829	11843	16207	44879
0.97–1	102438	65866	98096	266400	17217	12177	17116	46510
1–1.03	99652	63418	91491	254561	16897	11750	16430	45077
1.03–1.06	92177	59610	86991	238778	15815	11204	15664	42683
> 1.06	506405	445070	653051	1604526	97161	89247	124452	310860
Overall	1180854	980126	1438488	3599468	221941	191675	268393	682009

Table 4.1: Number of options binned by moneyness-maturity category for all 118 LIFFE option symbols and 25 LIFFE option symbols with non-dividend paying underlying.

	All 118 stocks				25 non-dividend paying stocks			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.31	1.05	2.60	1.49	0.26	1.01	1.82	1.11
0.94–0.97	0.79	2.10	4.11	2.35	0.61	1.75	2.58	1.62
0.97–1	1.24	2.64	4.63	2.84	0.89	2.12	2.98	1.98
1–1.03	1.85	3.23	5.01	3.33	1.20	2.41	3.11	2.21
1.03–1.06	2.59	3.96	5.67	4.05	1.58	2.97	3.42	2.62
> 1.06	6.46	7.95	8.98	7.90	3.64	5.26	4.80	4.57
Overall	3.38	4.68	6.08	4.81	1.98	3.30	3.49	2.95

Table 4.2: Average option price in Euros (€) in each moneyness-maturity bin for all 118 LIFFE option symbols and for 25 LIFFE option symbols with non dividend paying underlying.

	2009 SPX index options				2010 SPX index options		
	< 60	60–180	> 180	Overall	< 60	60–180	Overall
< 0.94	816	196	67	1079	3157	1484	4641
0.94–0.97	491	110	35	636	1928	293	2221
0.97–1	600	198	44	842	1951	405	2356
1–1.03	698	184	29	911	1678	314	1992
1.03–1.06	590	131	21	742	1119	90	1209
> 1.06	1164	300	9	1473	1717	231	1948
Overall	4359	1119	205	5683	11550	2817	14367

Table 4.3: Number of SPX options in 2009 and 2010 in each moneyness-maturity bin.

	2009 SPX index options				2010 SPX index options		
	< 60	60–180	> 180	Overall	< 60	60–180	Overall
< 0.94	7.73	29.49	52.02	14.44	2.88	9.70	5.06
0.94–0.97	20.98	53.30	81.42	29.90	8.80	33.03	12.00
0.97–1	31.65	66.06	97.47	43.18	20.43	50.13	25.53
1–1.03	43.50	74.77	110.95	51.96	37.94	64.81	42.18
1.03–1.06	57.26	89.43	127.96	64.94	59.51	87.86	61.62
> 1.06	116.77	138.94	160.66	121.55	122.25	146.97	125.18
Overall	54.06	82.11	87.68	60.80	35.16	37.84	35.68

Table 4.4: Average SPX index option price in US dollars ($) in each moneyness-maturity category in 2009 and 2010 SPX.

In-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0939	0.1767	0.3272	0.2164	0.0865	0.1484	0.2638	0.1793	0.0726	0.1263	0.2874	0.1789
0.94–0.97	0.1503	0.2204	0.3175	0.2300	0.1440	0.2114	0.2799	0.2114	0.1114	0.1358	0.2788	0.1798
0.97–1	0.1762	0.2325	0.3078	0.2386	0.1720	0.2311	0.2838	0.2278	0.1257	0.1411	0.2945	0.1917
1–1.03	0.1910	0.2563	0.3057	0.2485	0.1863	0.2533	0.2937	0.2416	0.1222	0.1473	0.3181	0.1989
1.03–1.06	0.1919	0.2810	0.3070	0.2561	0.1787	0.2664	0.2969	0.2437	0.1089	0.1547	0.3066	0.1924
>1.06	0.1440	0.3385	0.4005	0.3024	0.1236	0.2251	0.3296	0.2356	0.0631	0.1658	0.3653	0.2146
Overall	0.1431	0.2683	0.3559	0.2623	0.1302	0.2070	0.3000	0.2190	0.0834	0.1489	0.3258	0.1981

Out-of-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0961	0.1827	0.3340	0.2216	0.0887	0.1550	0.2711	0.1849	0.0777	0.1393	0.2991	0.1890
0.94–0.97	0.1532	0.2286	0.3255	0.2361	0.1468	0.2194	0.2881	0.2175	0.1195	0.1523	0.2900	0.1911
0.97–1	0.1791	0.2408	0.3158	0.2447	0.1750	0.2388	0.2924	0.2340	0.1352	0.1585	0.3074	0.2044
1–1.03	0.1928	0.2623	0.3150	0.2540	0.1882	0.2590	0.3023	0.2468	0.1318	0.1638	0.3285	0.2105
1.03–1.06	0.1929	0.2857	0.3162	0.2610	0.1802	0.2716	0.3065	0.2490	0.1182	0.1712	0.3189	0.2045
>1.06	0.1439	0.3396	0.4044	0.3042	0.1245	0.2297	0.3373	0.2403	0.0667	0.1747	0.3753	0.2222
Overall	0.1443	0.2723	0.3619	0.2661	0.1319	0.2127	0.3079	0.2242	0.0892	0.1611	0.3367	0.2076

Out-of-Sample Hedging Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0539	0.0554	0.0502	0.0527	0.0336	0.0391	0.0381	0.0371	0.0483	0.0623	0.0640	0.0591
0.94–0.97	0.0547	0.0518	0.0727	0.0607	0.0475	0.0484	0.0445	0.0466	0.0700	0.0836	0.0782	0.0765
0.97–1	0.0603	0.0705	0.0915	0.0744	0.0521	0.0521	0.0470	0.0502	0.0753	0.0888	0.0802	0.0805
1–1.03	0.1117	0.1016	0.1081	0.1079	0.0536	0.0528	0.0491	0.0517	0.0732	0.0889	0.0835	0.0809
1.03–1.06	0.1933	0.1444	0.1341	0.1593	0.0529	0.0557	0.0525	0.0534	0.0654	0.0897	0.0847	0.0786
>1.06	0.4178	0.3333	0.2485	0.3250	0.0378	0.0551	0.0562	0.0501	0.0366	0.0699	0.0772	0.0625
Overall	0.2278	0.1913	0.1537	0.1880	0.0415	0.0498	0.0490	0.0468	0.0510	0.0725	0.0747	0.0664

Table 4.5: All 118 LIFFE option symbols: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in Euros (€), respectively.

In-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0684	0.1324	0.1880	0.1358	0.0634	0.1078	0.1494	0.1114	0.0421	0.0600	0.0820	0.0636
0.94–0.97	0.0936	0.1318	0.1605	0.1279	0.0903	0.1208	0.1407	0.1165	0.0567	0.0592	0.0824	0.0666
0.97–1	0.1029	0.1301	0.1624	0.1319	0.1001	0.1246	0.1497	0.1248	0.0600	0.0574	0.0896	0.0702
1–1.03	0.1035	0.1330	0.1554	0.1301	0.0999	0.1276	0.1476	0.1245	0.0559	0.0556	0.0893	0.0680
1.03–1.06	0.1054	0.1470	0.1576	0.1355	0.0971	0.1363	0.1502	0.1269	0.0500	0.0556	0.0886	0.0656
>1.06	0.0843	0.1844	0.2030	0.1605	0.0700	0.1139	0.1606	0.1189	0.0320	0.0611	0.1028	0.0687
Overall	0.0852	0.1573	0.1879	0.1459	0.0763	0.1154	0.1540	0.1179	0.0418	0.0598	0.0930	0.0670

Out-of-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0705	0.1379	0.1934	0.1402	0.0656	0.1146	0.1555	0.1165	0.0471	0.0727	0.0919	0.0728
0.94–0.97	0.0962	0.1407	0.1661	0.1332	0.0929	0.1304	0.1470	0.1224	0.0643	0.0756	0.0933	0.0777
0.97–1	0.1055	0.1377	0.1682	0.1370	0.1028	0.1330	0.1561	0.1303	0.0680	0.0735	0.1010	0.0816
1–1.03	0.1059	0.1420	0.1630	0.1361	0.1024	0.1368	0.1547	0.1304	0.0647	0.0732	0.1002	0.0798
1.03–1.06	0.1067	0.1540	0.1644	0.1403	0.0990	0.1441	0.1572	0.1322	0.0588	0.0733	0.1001	0.0778
>1.06	0.0845	0.1867	0.2053	0.1622	0.0713	0.1199	0.1652	0.1228	0.0356	0.0712	0.1110	0.0760
Overall	0.0866	0.1620	0.1921	0.1493	0.0782	0.1223	0.1596	0.1226	0.0472	0.0723	0.1024	0.0759

Out-of-Sample Hedging Errors

	Black-Scholes				Markov Tree				Heston			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.0382	0.0408	0.0355	0.0378	0.0267	0.0330	0.0258	0.0281	0.0351	0.0508	0.0432	0.0430
0.94–0.97	0.0361	0.0454	0.0578	0.0465	0.0343	0.0374	0.0288	0.0331	0.0472	0.0623	0.0502	0.0523
0.97–1	0.0448	0.0628	0.0705	0.0591	0.0366	0.0392	0.0311	0.0353	0.0493	0.0632	0.0518	0.0539
1–1.03	0.0740	0.0821	0.0796	0.0782	0.0363	0.0368	0.0295	0.0339	0.0470	0.0610	0.0505	0.0520
1.03–1.06	0.1191	0.1142	0.0934	0.1083	0.0360	0.0410	0.0305	0.0353	0.0447	0.0639	0.0506	0.0520
>1.06	0.2657	0.2343	0.1485	0.2096	0.0266	0.0387	0.0318	0.0322	0.0260	0.0504	0.0464	0.0412
Overall	0.1476	0.1399	0.0980	0.1258	0.0294	0.0370	0.0296	0.0317	0.0348	0.0536	0.0466	0.0448

Table 4.6: 25 LIFFE option symbols with non-dividend paying underlying: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in Euros (€), respectively.

In-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.8066	1.2891	3.3626	1.0529	0.6531	1.1272	3.1615	0.8950	0.3827	0.4417	0.4571	0.3981
0.94–0.97	1.5875	1.5894	3.5900	1.6980	1.2043	1.2083	3.2579	1.3180	0.9695	0.7099	0.7592	0.9130
0.97–1	1.4227	1.4853	3.7589	1.5595	1.1894	1.1438	3.4583	1.2972	0.9834	0.7621	1.2557	0.9456
1–1.03	1.4231	1.5424	3.1430	1.5019	1.2005	1.2169	3.1522	1.2660	0.8087	0.7603	1.3450	0.8160
1.03–1.06	1.3544	1.8294	2.2189	1.4627	1.0814	1.3002	2.4501	1.1588	0.5964	0.6742	1.5565	0.6373
> 1.06	1.0156	1.3396	4.4884	1.1028	0.5414	0.9417	3.5248	0.6411	0.5892	0.4482	1.1743	0.5641
Overall	1.2080	1.4718	3.3877	1.3386	0.9048	1.1234	3.1834	1.0300	0.6838	0.6061	0.9498	0.6781

Out-of-Sample Pricing Errors

	Black-Scholes				Markov Tree				Heston			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	1.0283	1.8973	3.3559	1.3307	0.9923	1.9481	3.0750	1.2952	1.1567	2.4181	2.6902	1.4810
0.94–0.97	1.9542	2.3008	4.2466	2.1403	1.7276	2.3129	3.8562	1.9460	2.0058	2.5075	2.6868	2.1301
0.97–1	1.7818	2.1510	4.4395	2.0075	1.6928	2.1234	4.1764	1.9238	1.9189	2.2149	3.0285	2.0465
1–1.03	1.7268	1.8951	4.0414	1.8345	1.6345	1.7578	4.1126	1.7383	1.6763	1.9303	3.3505	1.7809
1.03–1.06	1.6237	2.1759	3.6915	1.7820	1.5654	1.8046	4.1257	1.6818	1.4313	1.8417	3.0522	1.5514
> 1.06	1.1800	1.6689	3.9170	1.2961	1.2352	1.5369	3.3974	1.3095	1.1188	1.4671	4.6104	1.2100
Overall	1.4704	1.9532	3.8964	1.6536	1.4174	1.8569	3.7149	1.5874	1.4694	1.9887	2.9690	1.6264

Out-of-Sample Hedging Errors

	Black-Scholes				Markov Tree				Heston			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	1.5960	1.6972	1.7254	1.6192	1.0208	1.5097	2.1006	1.1582	1.5538	3.0197	4.2792	1.9369
0.94–0.97	1.6073	1.4276	1.4424	1.5701	1.5660	1.8457	1.7572	1.6203	2.4214	3.8394	4.0458	2.7325
0.97–1	1.5668	1.4129	1.7418	1.5444	1.8276	1.7252	1.8811	1.8091	2.7939	3.4550	3.9352	2.9958
1–1.03	2.4576	1.8226	2.5395	2.3573	1.7119	1.8393	2.0883	1.7452	2.6189	3.6280	4.1940	2.8353
1.03–1.06	3.9381	2.6151	2.3725	3.7387	1.5337	1.4643	1.3571	1.5206	2.1383	2.7559	2.8632	2.2313
> 1.06	6.3086	4.5227	3.2818	5.9344	1.2601	1.2305	3.4344	1.2650	1.5567	1.8636	6.1118	1.6413
Overall	2.9351	2.3265	1.8930	2.7908	1.4548	1.5846	1.9519	1.4964	2.1303	3.0328	4.0640	2.3595

Table 4.7: SPX index options from 2009: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in US Dollars ($), respectively.

In-Sample Pricing Errors

	Black-Scholes			Markov Tree			Heston		
	< 60	60–180	Overall	< 60	60–180	Overall	< 60	60–180	Overall
< 0.94	2.4793	4.2638	3.0499	1.8093	3.1727	2.2453	0.7678	0.7232	0.7536
0.94–0.97	2.1529	1.3254	2.0437	1.5996	1.1367	1.5385	1.9971	1.3807	1.9158
0.97–1	1.5151	3.0341	1.7762	1.3373	2.8471	1.5968	3.3062	1.2439	2.9517
1–1.03	1.9801	5.6286	2.5552	1.6103	4.7825	2.1104	3.6362	1.2564	3.2611
1.03–1.06	2.9051	8.2283	3.3014	1.7680	6.3371	2.1081	3.3663	1.2077	3.2056
> 1.06	2.5464	9.5136	3.3726	1.8835	5.4483	2.3062	2.3409	1.6233	2.2558
Overall	2.2406	4.4907	2.6818	1.6727	3.3813	2.0077	2.3041	1.0152	2.0514

Out-of-Sample Pricing Errors

	Black-Scholes			Markov Tree			Heston		
	< 60	60–180	Overall	< 60	60–180	Overall	< 60	60–180	Overall
< 0.94	2.5247	4.6013	3.1887	2.0081	3.9160	2.6181	1.1167	1.6214	1.2780
0.94–0.97	2.5631	3.5518	2.6935	2.1082	3.3628	2.2737	2.5404	3.0072	2.6020
0.97–1	2.2363	4.2364	2.5801	2.0710	3.9689	2.3973	3.9287	2.7678	3.7291
1–1.03	2.3524	5.6459	2.8715	2.0468	4.6468	2.4566	4.2024	2.5562	3.9429
1.03–1.06	2.9328	9.0714	3.3898	1.9820	6.4140	2.3119	3.7052	2.7069	3.6309
> 1.06	2.5203	9.7335	3.3761	2.1347	5.8657	2.5774	2.4940	2.9696	2.5504
Overall	2.4963	5.1198	3.0107	2.0573	4.1872	2.4750	2.7332	2.1798	2.6247

Out-of-Sample Hedging Errors

	Black-Scholes			Markov Tree			Heston		
	< 60	60–180	Overall	< 60	60–180	Overall	< 60	60–180	Overall
< 0.94	1.6121	1.9310	1.7063	0.8483	1.2550	0.9684	1.1571	1.8352	1.3574
0.94–0.97	1.4957	1.5417	1.5005	1.0896	1.7307	1.1570	2.0851	3.6284	2.2474
0.97–1	1.2584	1.3563	1.2722	1.3459	1.9216	1.4272	2.7414	4.1820	2.9447
1–1.03	1.9874	1.4087	1.9101	1.4541	1.7795	1.4976	2.6213	3.6391	2.7572
1.03–1.06	3.9285	2.6833	3.8763	1.5916	1.7519	1.5983	2.3154	3.6511	2.3714
> 1.06	6.5603	4.8238	6.4119	1.6083	2.7452	1.7054	1.9719	4.2644	2.1679
Overall	2.2491	1.8759	2.1850	1.2223	1.5396	1.2768	2.0465	2.7290	2.1638

Table 4.8: SPX index options from 2010: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in US Dollars ($), respectively.

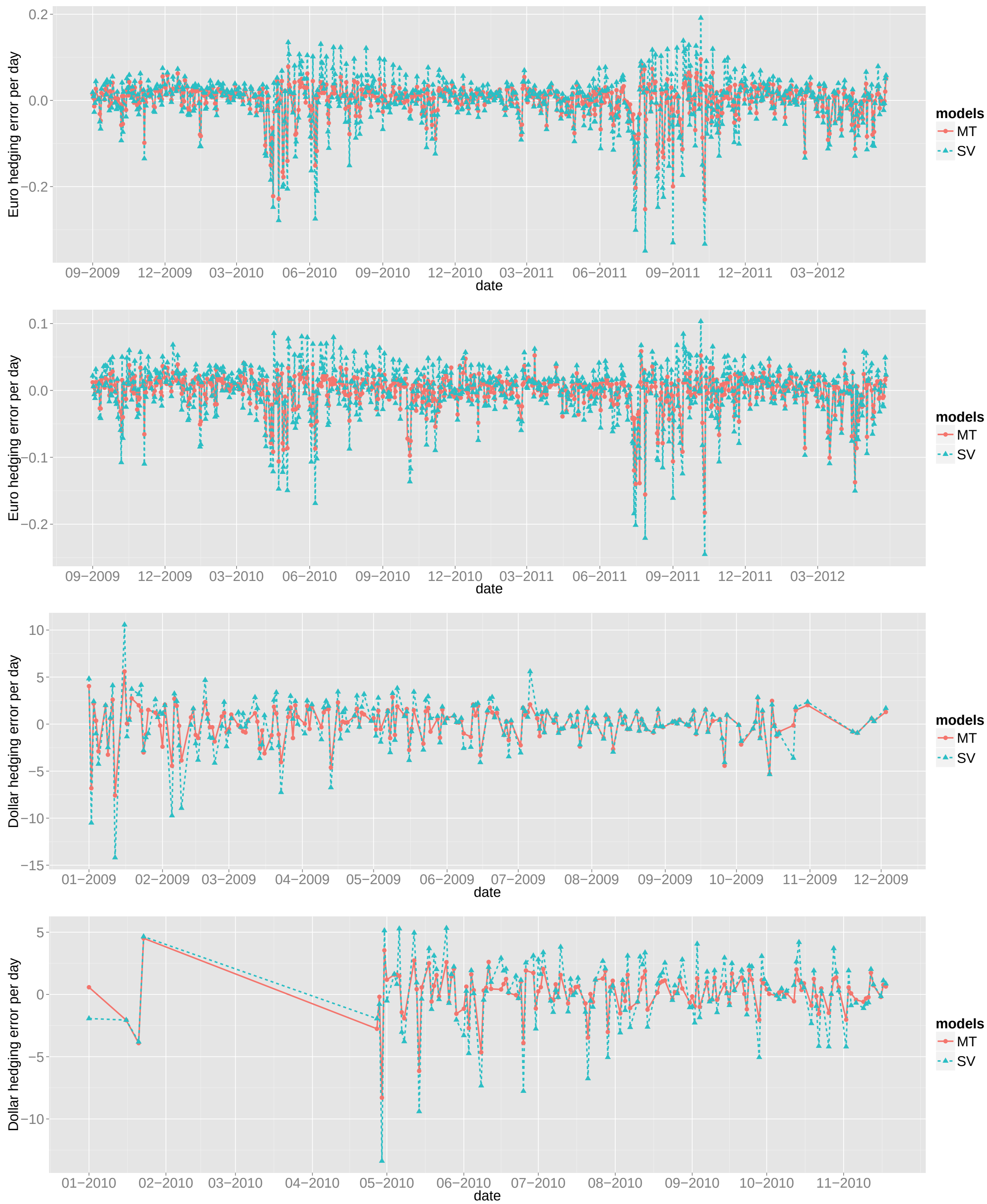

Figure 4.1: We plot the daily market hedging error against the corresponding date. In the first and second panel, we plot the market hedging errors in Euros (€) for all 118 LIFFE option symbols and 25 LIFFE option symbols with non dividend paying underlying respectively. In the bottom two panels, we plot the market hedging errors in US Dollars ($) for SPX index options from 2009 and 2010 respectively.

In-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0865	0.1484	0.2638	0.1793	0.0735	0.1219	0.2726	0.1716				
0.94–0.97	0.1440	0.2114	0.2799	0.2114	0.1245	0.1723	0.2801	0.1943				
0.97–1	0.1720	0.2311	0.2838	0.2278	0.1498	0.1908	0.2851	0.2098				
1–1.03	0.1863	0.2533	0.2937	0.2416	0.1634	0.2135	0.2957	0.2235				
1.03–1.06	0.1787	0.2664	0.2969	0.2437	0.1581	0.2291	0.2980	0.2268				
> 1.06	0.1236	0.2251	0.3296	0.2356	0.1097	0.2008	0.3440	0.2303				
Overall	0.1302	0.2070	0.3000	0.2190	0.1141	0.1782	0.3094	0.2096				

Out-of-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0887	0.1550	0.2711	0.1849	0.0769	0.1316	0.2826	0.1796	0.0242	0.0335	0.0330	0.0306
0.94–0.97	0.1468	0.2194	0.2881	0.2175	0.1286	0.1838	0.2904	0.2026	0.0322	0.0417	0.0395	0.0373
0.97–1	0.1750	0.2388	0.2924	0.2340	0.1540	0.2011	0.2963	0.2180	0.0355	0.0448	0.0422	0.0403
1–1.03	0.1882	0.2590	0.3023	0.2468	0.1662	0.2214	0.3071	0.2306	0.0347	0.0452	0.0442	0.0407
1.03–1.06	0.1802	0.2716	0.3065	0.2490	0.1600	0.2359	0.3107	0.2338	0.0317	0.0449	0.0449	0.0398
> 1.06	0.1245	0.2297	0.3373	0.2403	0.1110	0.2067	0.3541	0.2365	0.0239	0.0437	0.0694	0.0478
Overall	0.1319	0.2127	0.3079	0.2242	0.1165	0.1860	0.3198	0.2166	0.0272	0.0409	0.0521	0.0409

Out-of-Sample Hedging Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0336	0.0391	0.0381	0.0371	0.0332	0.0392	0.0392	0.0375	0.0346	0.0389	0.0360	0.0364
0.94–0.97	0.0475	0.0484	0.0445	0.0466	0.0474	0.0486	0.0455	0.0470	0.0449	0.0469	0.0420	0.0443
0.97–1	0.0521	0.0521	0.0470	0.0502	0.0522	0.0523	0.0478	0.0506	0.0483	0.0500	0.0443	0.0472
1–1.03	0.0536	0.0528	0.0491	0.0517	0.0533	0.0529	0.0497	0.0519	0.0476	0.0507	0.0457	0.0477
1.03–1.06	0.0529	0.0557	0.0525	0.0534	0.0520	0.0554	0.0527	0.0531	0.0431	0.0494	0.0455	0.0455
> 1.06	0.0378	0.0551	0.0562	0.0501	0.0369	0.0542	0.0556	0.0494	0.0284	0.0405	0.0459	0.0388
Overall	0.0415	0.0498	0.0490	0.0468	0.0410	0.0495	0.0492	0.0466	0.0358	0.0423	0.0428	0.0403

Table 4.9: Comparison of three different regression procedures for the Markov tree model on 118 LIFFE option symbols: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in euros (€), respectively.

In-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0634	0.1078	0.1494	0.1114	0.0545	0.0882	0.1569	0.1061				
0.94–0.97	0.0903	0.1208	0.1407	0.1165	0.0775	0.0963	0.1427	0.1060				
0.97–1	0.1001	0.1246	0.1497	0.1248	0.0860	0.0985	0.1567	0.1153				
1–1.03	0.0999	0.1276	0.1476	0.1245	0.0858	0.1033	0.1515	0.1143				
1.03–1.06	0.0971	0.1363	0.1502	0.1269	0.0840	0.1109	0.1532	0.1165				
> 1.06	0.0700	0.1139	0.1606	0.1189	0.0607	0.0961	0.1668	0.1133				
Overall	0.0763	0.1154	0.1540	0.1179	0.0659	0.0953	0.1601	0.1112				

Out-of-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0656	0.1146	0.1555	0.1165	0.0574	0.0966	0.1647	0.1126	0.0186	0.0266	0.0220	0.0223
0.94–0.97	0.0929	0.1304	0.1470	0.1224	0.0810	0.1083	0.1491	0.1128	0.0225	0.0299	0.0247	0.0252
0.97–1	0.1028	0.1330	0.1561	0.1303	0.0896	0.1089	0.1637	0.1219	0.0244	0.0311	0.0251	0.0264
1–1.03	0.1024	0.1368	0.1547	0.1304	0.0894	0.1143	0.1594	0.1214	0.0236	0.0320	0.0252	0.0264
1.03–1.06	0.0990	0.1441	0.1572	0.1322	0.0864	0.1208	0.1612	0.1229	0.0220	0.0314	0.0251	0.0256
> 1.06	0.0713	0.1199	0.1652	0.1228	0.0625	0.1035	0.1724	0.1182	0.0161	0.0264	0.0238	0.0221
Overall	0.0782	0.1223	0.1596	0.1226	0.0684	0.1038	0.1667	0.1170	0.0189	0.0276	0.0236	0.0232

Out-of-Sample Hedging Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall	< 60	60–180	> 180	Overall
< 0.94	0.0267	0.0330	0.0258	0.0281	0.0264	0.0330	0.0265	0.0283	0.0265	0.0315	0.0241	0.0270
0.94–0.97	0.0343	0.0374	0.0288	0.0331	0.0342	0.0374	0.0293	0.0333	0.0318	0.0355	0.0269	0.0310
0.97–1	0.0366	0.0392	0.0311	0.0353	0.0366	0.0393	0.0316	0.0354	0.0334	0.0356	0.0271	0.0316
1–1.03	0.0363	0.0368	0.0295	0.0339	0.0361	0.0367	0.0297	0.0339	0.0336	0.0381	0.0273	0.0324
1.03–1.06	0.0360	0.0410	0.0305	0.0353	0.0355	0.0408	0.0307	0.0352	0.0309	0.0359	0.0265	0.0306
> 1.06	0.0266	0.0387	0.0318	0.0322	0.0259	0.0380	0.0314	0.0316	0.0205	0.0294	0.0250	0.0248
Overall	0.0294	0.0370	0.0296	0.0317	0.0290	0.0367	0.0297	0.0315	0.0257	0.0317	0.0252	0.0272

Table 4.10: Comparison of three different regression procedures for the Markov tree model on 25 LIFFE option symbols with non-dividend paying underlying: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in euros (€), respectively.

In-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.6531	1.1272	3.1615	0.8950	0.6076	1.1512	3.7982	0.9045				
0.94–0.97	1.2043	1.2083	3.2579	1.3180	1.1659	1.1951	3.7680	1.3141				
0.97–1	1.1894	1.1438	3.4583	1.2972	1.1216	1.1430	4.1478	1.2848				
1–1.03	1.2005	1.2169	3.1522	1.2660	1.0907	1.1261	3.9542	1.1890				
1.03–1.06	1.0814	1.3002	2.4501	1.1588	0.9630	1.2113	3.1659	1.0692				
>1.06	0.5414	0.9417	3.5248	0.6411	0.4576	1.0254	5.8086	0.6059				
Overall	0.9048	1.1234	3.1834	1.0300	0.8267	1.1233	3.9136	0.9964				

Out-of-Sample Pricing Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	0.9923	1.9481	3.0750	1.2952	0.9699	1.9284	3.5258	1.3027	0.7289	1.4535	1.8470	0.9067
0.94–0.97	1.7276	2.3129	3.8562	1.9460	1.7557	2.3093	4.4246	1.9983	1.1513	1.5673	1.7941	1.2470
0.97–1	1.6928	2.1234	4.1764	1.9238	1.7074	2.1721	4.8128	1.9790	1.1478	1.3820	1.3632	1.2071
1–1.03	1.6345	1.7578	4.1126	1.7383	1.6111	1.7243	4.8980	1.7389	1.1457	1.2469	1.6519	1.1784
1.03–1.06	1.5654	1.8046	4.1257	1.6818	1.5343	1.9275	4.8187	1.7028	1.0785	1.3287	1.3939	1.1183
>1.06	1.2352	1.5369	3.3974	1.3095	1.1814	1.6359	3.7108	1.2851	0.9005	0.9412	2.4268	0.9175
Overall	1.4174	1.8569	3.7149	1.5874	1.3978	1.8976	4.3047	1.6019	1.0014	1.2913	1.6722	1.0756

Out-of-Sample Hedging Errors

	Overconstrained L^2				Pseudo-Huber				Underconstrained L^2			
	<60	60–180	>180	Overall	<60	60–180	>180	Overall	<60	60–180	>180	Overall
<0.94	1.0208	1.5097	2.1006	1.1582	1.0213	1.5106	2.1209	1.1599	0.9814	1.5389	2.0052	1.1174
0.94–0.97	1.5660	1.8457	1.7572	1.6203	1.5755	1.8647	1.7848	1.6323	1.6725	1.7379	1.8819	1.6918
0.97–1	1.8276	1.7252	1.8811	1.8091	1.8346	1.7370	1.8993	1.8177	1.6571	1.4953	1.5277	1.6211
1–1.03	1.7119	1.8393	2.0883	1.7452	1.7155	1.8535	2.1170	1.7514	1.5463	1.2259	1.3636	1.4969
1.03–1.06	1.5337	1.4643	1.3571	1.5206	1.5061	1.4700	1.3488	1.4974	1.2811	1.3652	1.0848	1.2846
>1.06	1.2601	1.2305	3.4344	1.2650	1.1600	1.2270	2.3251	1.1784	0.9971	1.0844	1.1629	1.0126
Overall	1.4548	1.5846	1.9519	1.4964	1.4372	1.5916	1.9423	1.4835	1.3278	1.4058	1.6783	1.3524

Table 4.11: Comparison of three different regression procedures for the Markov tree model on 2009 SPX index options: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in US Dollars ($), respectively.

In-Sample Pricing Errors

	Overconstrained L^2			Pseudo-Huber			Underconstrained L^2		
	<60	60–180	Overall	<60	60–180	Overall	<60	60–180	Overall
<0.94	1.8093	3.1727	2.2453	1.5200	2.6941	1.8954			
0.94–0.97	1.5996	1.1367	1.5385	1.1958	1.5735	1.2456			
0.97–1	1.3373	2.8471	1.5968	1.2525	3.7508	1.6820			
1–1.03	1.6103	4.7825	2.1104	1.8815	5.8175	2.5019			
1.03–1.06	1.7680	6.3371	2.1081	2.0258	7.2277	2.4130			
>1.06	1.8835	5.4483	2.3062	1.9338	6.0313	2.4197			
Overall	1.6727	3.3813	2.0077	1.5837	3.4962	1.9587			

Out-of-Sample Pricing Errors

	Overconstrained L^2			Pseudo-Huber			Underconstrained L^2		
	<60	60–180	Overall	<60	60–180	Overall	<60	60–180	Overall
<0.94	2.0081	3.9160	2.6181	1.7537	3.5422	2.3256	0.4769	0.8814	0.5937
0.94–0.97	2.1082	3.3628	2.2737	1.8129	3.6109	2.0501	0.7766	1.4730	0.8478
0.97–1	2.0710	3.9689	2.3973	2.0111	4.7730	2.4859	1.0356	1.5060	1.1002
1–1.03	2.0468	4.6468	2.4566	2.2669	5.5526	2.7848	1.1075	1.3741	1.1403
1.03–1.06	1.9820	6.4140	2.3119	2.2277	7.2449	2.6012	1.0671	1.4748	1.0859
>1.06	2.1347	5.8657	2.5774	2.2293	6.2374	2.7042	1.0882	1.9201	1.1545
Overall	2.0573	4.1872	2.4750	1.9981	4.2890	2.4474	0.8511	1.1511	0.9009

Out-of-Sample Hedging Errors

	Overconstrained L^2			Pseudo-Huber			Underconstrained L^2		
	<60	60–180	Overall	<60	60–180	Overall	<60	60–180	Overall
<0.94	0.8483	1.2550	0.9684	0.8024	1.2060	0.9216	0.6159	0.9517	0.7065
0.94–0.97	1.0896	1.7307	1.1570	1.0639	1.7000	1.1308	0.9466	1.4735	0.9919
0.97–1	1.3459	1.9216	1.4272	1.3324	1.9071	1.4135	1.1796	1.5360	1.2227
1–1.03	1.4541	1.7795	1.4976	1.4537	1.7812	1.4975	1.1644	1.2645	1.1742
1.03–1.06	1.5916	1.7519	1.5983	1.6123	1.7623	1.6186	1.0635	1.2231	1.0686
>1.06	1.6083	2.7452	1.7054	1.6270	2.7796	1.7256	1.0783	1.8883	1.1236
Overall	1.2223	1.5396	1.2768	1.2055	1.5083	1.2575	0.9572	1.1610	0.9872

Table 4.12: Comparison of three different regression procedures for the Markov tree model on 2010 SPX index options: from top to bottom, we present the in-sample, one day out-of-sample, and out-of-sample hedging mean absolute erorrs in US Dollars ($), respectively.

	λ	$\bar{\alpha}$	μ	σ	γ
2009	-0.4022	0.5242	-0.0803	1.4903	0.0980
2010	1.2218	0.3808	-1.8850	2.442	1.4059

Table 4.13: Best fit GHD parameters on errors from 2009 and 2010.

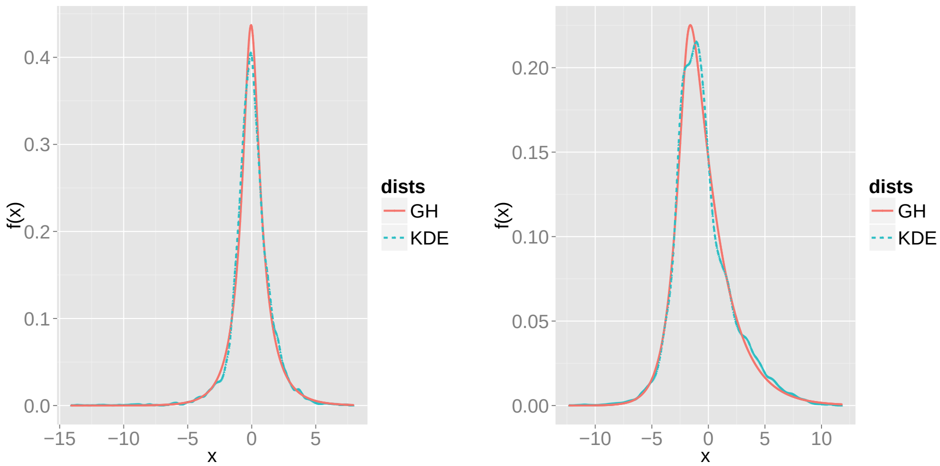

Figure 4.2: 2009 and 2010 in-sample error distribution using KDE and best fit GHD

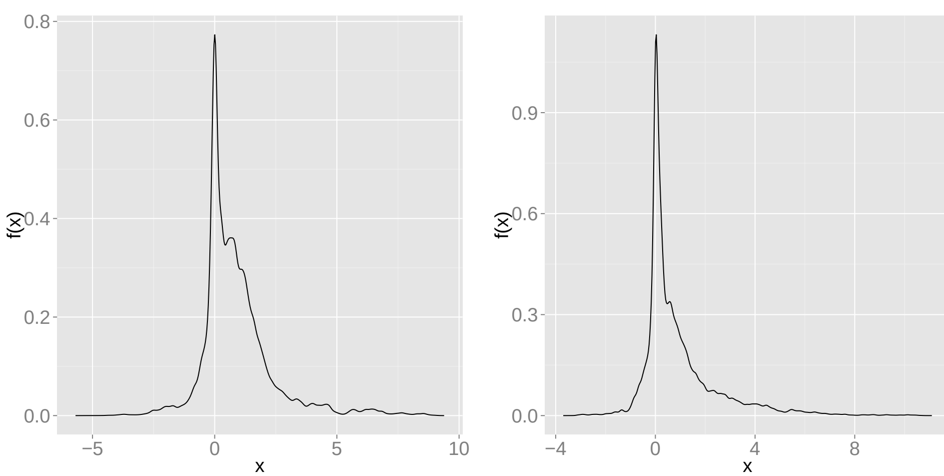

Figure 4.3: KDE of 2009 and 2010 $|H^{\mathrm{SV}}| - |H^{\mathrm{MT}}|$

	10%	20%	30%	40%	50%	60%	70%	80%	90%
2009	-0.2888	-0.0357	0.0651	0.2879	0.5720	0.8463	1.1781	1.5894	2.4213
2010	-0.2008	-0.0066	0.0578	0.1768	0.3679	0.6658	1.0401	1.6276	2.8967

Table 4.14: Deciles of $|H^{\mathrm{SV}}| - |H^{\mathrm{MT}}|$ for 2009 and 2010.

Chapter 5

Generalization of the Markov Tree Process

Noise and time delays are key features of models of human balance (Ohira and Milton, 1995; Milton, 2011), circadian oscillators (Smolen et al., 2002), gene regulation dynamics (Bratsun et al., 2005; Josic et al., 2011), cortical interneuron migration (Tanaka et al., 2009), and resting brain dynamics (Deco et al., 2009, 2010). Despite the success of spectral methods in other stochastic contexts (Bhattacharya and Waymire, 2009; Mugler et al., 2009), delayed stochastic systems are typically not treated using spectral methods (Longtin, 2010).

In this Chapter, we present a spectral numerical method to calculate the probability density function (pdf) for the delayed random walk that is obtained by applying the weak Euler-Maruyama discretization to a class of stochastic delay differential equations (SDDE). We refer to the method as a spectral method because it involves solving the problem in Fourier space, and then using the inverse FFT (fast Fourier transform) to compute the solution in physical space. This method is fast, exact (to machine precision) and generalizable to other, more complicated systems.

5.1 Introduction

Consider the SDDE

$$dY_t = \phi(Y_t - Y_{t-\ell dt})dt + \gamma(Y_t - Y_{t-\ell dt})dW_t \tag{5.1}$$

with initial conditions $Y(t) = \theta(t) \quad$ for $t \in [0, \ell dt]$. Here ℓ is the integer delay (lag), W_t is the standard Wiener process, and ϕ and γ are measurable functions subject to the condition that when $\gamma \equiv 0$, the resulting deterministic equation has a stable fixed point.

To obtain the pdf of a stochastic differential equation (with no delay) at time $t > 0$, a natural approach is to solve the associated Fokker-Planck equation. For an SDDE, however, the delayed Fokker-Planck equation is circular (Longtin, 2010) and cannot be solved using standard numerical methods (Frank, 2005b). For this reason, past studies have applied asymptotic and perturbative methods to extract useful information from delayed Fokker-Planck equations (Guillouzic et al., 1999; Frank, 2005b,a; Galla, 2009). Such methods break down when the noise term is multiplied by a function of the delayed solution, or when the delay is large.

The technique employed in this Chapter is fundamentally different from the Fokker-Planck approach. We use a standard stochastic numerical method to discretize (5.1) in time and space. This discretization, together with piecewise constant approximation of the functions ϕ and γ, yields a delayed random walk approximation (5.4) of the original SDDE, the pdf of which is then computed using a fast and accurate spectral method.

An important reason for taking this approach is that delayed random walks can often be solved exactly (Ohira and Milton, 2009). Prior delayed random walk approximations to SDDE (Ohira and Yamane, 2000; Ohira and Milton, 2009) feature a V-shaped potential such that the walker's probabilities of right/left movements are spatially dependent. The equivalence between the Fokker-Planck equations for this delayed random walk and the original SDDE has been demonstrated (Ohira and Milton, 2009), generalizing ideas of Ehrenfest and Kac.

Instead of using a spatially dependent potential, the delayed random walk approximation (5.4) allows for non-uniformity of both the sizes $K_r^{\pm}$ and probabilities $\{q_r, 1-q_r\}$ of the increments; through r, these quantities also have a piecewise constant dependence on space. This (piecewise) spatial homogeneity allows us to rewrite the system as a recursion that can be solved using spatial Fourier transforms. We view (5.6) as a discrete equation for the approximate time-evolution of the pdf of (5.1). Note that (5.6) differs both in derivation and solution from Fokker-Planck equations for SDDE (Guillouzic et al., 1999; Frank, 2005b,a; Galla, 2009).

5.2 Delayed Random Walk

We discretize SDDE (5.1) using the weak Euler-Maruyama scheme (Higham, 2001) to obtain

$$Y_{n+1} = Y_n + \phi(Y_n - Y_{n-\ell})\Delta t + \gamma(Y_n - Y_{n-\ell})\sqrt{\Delta t}Z, \tag{5.2}$$

where Z is a Bernoulli random variable that takes values $\{-1, 1\}$ with equal probabilities. The initial conditions given after (5.1) yield initial conditions for (5.2): $Y_j = \theta(jdt)$ for $j = 0, 1, \ldots, \ell$. Let $I_{\mathcal{A}}$ denote the indicator function on the set $\mathcal{A}$. We use

$$\phi(x) \approx \sum_{r=1}^{R} \mu_r I_{[c_r, c_{r+1})}(x), \quad \gamma(x) \approx \sum_{r=1}^{R} \sigma_r I_{[c_r, c_{r+1})}(x),$$

piecewise constant approximations with constant μ_r and σ_r, and substitute back into (5.2) to obtain

$$Y_{n+1} = Y_n + \mu_r \Delta t + \sigma_r \sqrt{\Delta t} Z,\ c_r \le Y_n - Y_{n-\ell} < c_{r+1}. \tag{5.3}$$

We rewrite (5.3) as the delayed random walk

$$\begin{aligned} Y_{n+1} &= Y_n + K_n, \\ K_n &= K_n^r \quad \text{if} \quad c_r \le Y_n - Y_{n-\ell} < c_{r+1}, \end{aligned} \tag{5.4}$$

where K_n^r is a Bernoulli random variable that takes values $\{K_r^+, K_r^-\}$ with probabilities $\{q_r, 1-q_r\}$ respectively. We choose $\{K_r^+, K_r^-, q_r\}$ such that the moments of K_n^r match those of $\mu_r\Delta t + \sigma_r\sqrt{\Delta t}Z$ [1]. The delayed random walk (5.4) has not been considered in the literature, to the best of our knowledge. This random walk is more general than exactly solvable delayed and/or persistent random walks in the literature (Berrones and Larralde, 2001; Weiss, 2002; Rudnick and Gaspari, 2004; Van der Straeten and Naudts, 2006; García-Pelayo, 2007; Bhat and Kumar, 2012).

[1] For the purposes of approximating the weak EM scheme of the SDDE, we set $K_n^r = \mu_r \Delta t + \sigma_r \sqrt{\Delta t} Z$.

5.3 Spectral Method

Let $\Omega = \{K_r^+, K_r^-\}_{r=1}^R$ and let α_j be the outcome of the random variable K_{n-j+1}. Applying Bayes' theorem recursively to (5.4), we get

$$\begin{aligned} &P(Y_{n+1} = s \cap K_n = \alpha_1 \cap \cdots \cap K_{n-\ell+1} = \alpha_\ell) \\ &\qquad = \sum_{\alpha_{\ell+1}\in\Omega} P(Y_n = s - \alpha_1 \cap K_{n-1} = \alpha_2 \cap \cdots \cap K_{n-\ell} = \alpha_{\ell+1}) \\ &\qquad\qquad \times P(K_n = \alpha_1 | K_{n-1} = \alpha_2 \cap \cdots \cap K_{n-\ell} = \alpha_{\ell+1}). \end{aligned} \tag{5.5}$$

Denote the left-hand side as $T_s^{n+1}(\alpha_1^\ell)$ and the conditional probability as $p(\alpha_1^{\ell+1})$. Then

$$T_s^{n+1}(\alpha_1^\ell) = \sum_{\alpha_{\ell+1}\in\Omega} T_{s-\alpha_1}^n(\alpha_2^{\ell+1})p(\alpha_1^{\ell+1}). \tag{5.6}$$

Taking the Fourier transform in s yields the linear system

$$\hat{T}_k^{n+1}(\alpha_1^\ell) = \sum_{\alpha_{\ell+1}\in\Omega} \hat{T}_k^n(\alpha_2^{\ell+1}) \underbrace{p(\alpha_1^{\ell+1})e^{-i2\pi k\alpha_1}}_{M} \tag{5.7}$$

in k space, where $\hat{T}_k^{n+1}(\alpha_1^\ell)$ denotes the Fourier transform of the probability of reaching s by taking a sequence of steps $\alpha_\ell, \ldots, \alpha_1$ in the previous ℓ steps. In (5.7), we use M to denote the $(2R)^\ell \times (2R)^\ell$ matrix that gives the probability of transitioning from a sequence of states $(\alpha_{\ell+1}, \ldots, \alpha_2)$ to the sequence $(\alpha_\ell, \ldots, \alpha_1)$. Sparsity of M follows easily: since each Bernoulli random variable has only two outcomes, there are exactly two non-zero entries in every column of M for a total of $2 \times (2R)^\ell$ non-zero entries. From (5.7) we have $\hat{v}^{n+1} = M\hat{v}^n$, which implies $\hat{v}_n = M^{n-2\ell}\hat{v}_{2\ell}$, where $\hat{v}_n$ is a $(2R)^\ell \times 1$ vector with each component representing $\hat{T}_k^n(\alpha_\ell, \ldots, \alpha_1)$. Let $f(n, s)$ denote the pdf of the delayed random walk (5.4) at time step n, and let $\hat{f}(n, k)$ denote its Fourier transform with k as the variable that is Fourier conjugate to s. Then, based on the above, we have derived the solution in Fourier space:

$$\hat{f}(n, k) = \mathbf{1}^T M^{n-2\ell}\hat{v}_{2\ell}. \tag{5.8}$$

To compute the initial condition $\hat{v}_{2\ell}$, we require two steps. First, we use the initial condition $Y_0, \ldots, Y_\ell$ in the modified tree method (described below) to compute the exact pdf of $Y_{2\ell}$, the solution of (5.4) at time $n = 2\ell$. Next, we set $\hat{v}_{2\ell}$ equal to the Fourier transform in space of the pdf of $Y_{2\ell}$. In this way, the spectral method handles any initial conditions $\{Y_j\}_{0\le j\le\ell}$ consisting of discrete random variables. This includes, for example, any set of constant initial conditions for (5.4), and therefore any piecewise constant initial function $\theta(t)$ for (5.1).

What remains is to recover $f(n, s)$ from $\hat{f}(n, k)$. Since the walk is discrete in space, $f(n, s)$ is a linear combination of Dirac delta functions,

$$f(n, s) = \sum_{m\in\mathcal{N}} f_m\delta(s - s_m), \tag{5.9}$$

where s_m takes specific values in s space depending on the parameters in the set Ω and $\mathcal{N} = \{-N/2, -N/2+1, \ldots, N/2-1\}$. The presence of the Dirac deltas is a reason to avoid naïve Fourier inversion of $\hat{f}(n,k)$. Here, note that f is determined completely by the set $\{(f_m, s_m)\}_{m\in\mathcal{N}}$: it is this set we will solve for.

With f represented by (5.9), its Fourier transform is $\hat{f}(n,k) = \sum_{m\in\mathcal{N}} f_m e^{-i2\pi k s_m}$. We sample $\hat{f}(n,k)$ at discrete values of k given by $k_j = j\Delta k$ for all $j \in \mathcal{N}$:

$$\hat{f}(n,k_j) = \sum_{m\in\mathcal{N}} f_m e^{-i2\pi j\Delta k s_m}. \tag{5.10}$$

Let $\hat{\delta}$ denote the Kronecker delta, and assume that $\Delta k \Delta s = 1/N$. Then the inverse FFT of (5.10) is

$$\begin{aligned} f(n,s_r) &= \frac{1}{N}\sum_{j\in\mathcal{N}}\sum_{m\in\mathcal{N}} f_m e^{-i2\pi j\Delta k s_m} e^{i2\pi j\Delta k s_r} \\ &= \frac{1}{N}\sum_{m\in\mathcal{N}} f_m \sum_{j\in\mathcal{N}} e^{i2\pi j\Delta k(r-m)\Delta s} \\ &= \frac{1}{N}\sum_{m\in\mathcal{N}} f_m N\hat{\delta}(r-m) = f_r. \end{aligned}$$

The spectral method can now be summarized. In the first step, we compute (5.8), the exact solution in Fourier space, but sampled only at discrete values of k given by $k_j = j\Delta k$ for all $j \in \mathcal{N}$. In the second step, we compute the IFFT of this sampled Fourier transform at all s_m such that $m \in \mathcal{N}$. As shown, this yields the exact weight f_m corresponding to the spatial location s_m, meaning that we can indeed recover the set $\{(f_m, s_m)\}_{m\in\mathcal{N}}$ that determines (5.9) exactly. We denote the solution produced by the spectral method as $f_{\mathrm{IFFT}}(n,s)$. The only source of error between $f_{\mathrm{IFFT}}(n,s)$ and the exact pdf $f(n,s)$ is due to the inaccuracy in the IFFT algorithm itself (Briggs and Henson, 1995).

Note that the first step requires computing the matrix-vector product n times to obtain the Fourier transform at N different points in k space, while the second step consists entirely of the IFFT. The total complexity of the spectral method is thus $N(2R)^\ell n + N\log N \sim O(n^2)$, lower than the tree-based method described below.

Choosing Δs and Δk. Since the parameters in Ω are not necessarily equal, we have a pdf over s space with non-uniform spacing. We first convert this non-uniform grid into a uniform grid in order to use the IFFT. Let $\{K_r^\pm\}_{r=1}^R$ be rationals such that L is the least common multiple (LCM) of their denominators. Since the random walker changes its position by an element of $\{K_r^\pm\}_{r=1}^R$ at every step, the minimum non-zero distance between two sites that the random walker can occupy is given by $\Delta s = 1/L$. The maximum and minimum s values that can be reached by the random walker at any step n are, respectively, $S_{\max} = n\max\{K_r^\pm\}_{r=1}^R$ and $S_{\min} = n\min\{K_r^\pm\}_{r=1}^R$. This also implies that we have to calculate the pdf at $N = (S_{\max} - S_{\min})/\Delta s \sim O(nL)$ number of grid points. Since L is a constant given the parameters, we get $N \sim O(n)$, where n is the number of steps taken by the random walker. Finally, using $\Delta k\Delta s = 1/N$, we get $\Delta k = 1/(N\Delta s) = L/N$. Note that the parameters in the set Ω can be approximated such that L is small. This leads to incurring a relatively small error in calculating the pdf, while increasing the efficiency of the

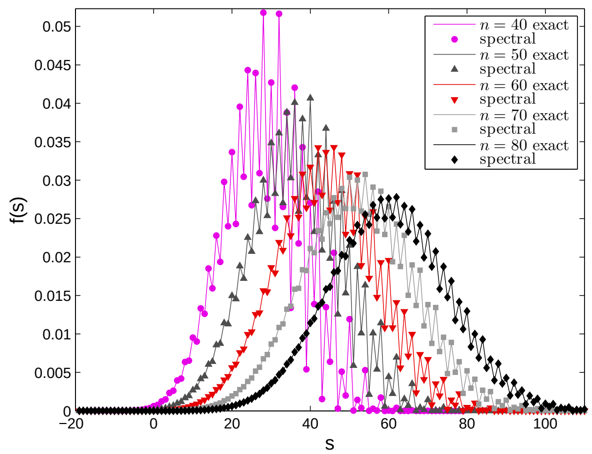

Figure 5.1: Snapshots at different values of time n show machine precision agreement between the densities computed using the spectral method (each plotted with a different marker) and an enumerative exact method (each plotted using the same grayscale/color as the corresponding marker). Computed densities are for the random walk (5.4) with delay $\ell = 5$ and two types of Bernoulli steps K_n: outcomes $\{2, -2\}$ with probabilities $\{0.7, 0.3\}$ when $Y_n \geq Y_{n-5}$, and outcomes $\{1, -1\}$ with probabilities $\{0.9, 0.1\}$ when $Y_n < Y_{n-5}$. Initial conditions were $Y_n = 0$ for $n \leq \ell$.

algorithm.

5.4 Modified Tree Method

For the delayed random walk (5.4), we have also developed an enumerative method for computing the exact pdf. This modified tree method involves growing a tree of all allowed paths/probabilities of the random walker. In previous work (Bhat and Kumar, 2012), the authors explained how to do this when $\ell = 1$. For $\ell > 1$, we modify the old procedure, leveraging the rationality of the increments of (5.4). Given the pdf at any step m consisting of $O(m)$ distinct states, computing the pdf at step $m + 1$ using the tree method requires three steps: (i) calculating all possible states at step $m + 1$, (ii) tracking the history and the region in which each of these states lie, and (iii) checking for recombinations to obtain the pdf at step $m+1$. Step (i) requires $2m$ operations, while (ii) requires $(2R)^\ell$ operations per state. Step (iii) requires finding unique states with the same history and summing the probabilities in each of these unique states. The overall complexity is then $\sum_{m=1}^{n}(2m)(2R)^\ell + m^2(2R)^\ell \sim O(n^3)$. In this work, we use this method for two purposes: to compute $\hat{v}_{2\ell}$ for (5.8), and to compute exact reference solutions against which we compare the spectral method.

5.5 Results

For both Fig. 5.1 and Fig. 5.2, we plot in solid lines (respectively, solid markers in the same grayscale/color) the pdf calculated by growing the tree (respectively, the spectral method). In these figures, different grayscales/colors and markers are used for different values of n and ℓ, respectively. The solid markers lie exactly on the solid curves, demonstrating the accuracy of the spectral method. In Fig. 5.3, we plot $\|f_{\mathrm{IFFT}}(n, s) - f(n, s)\|_\infty$ both for different numbers of steps

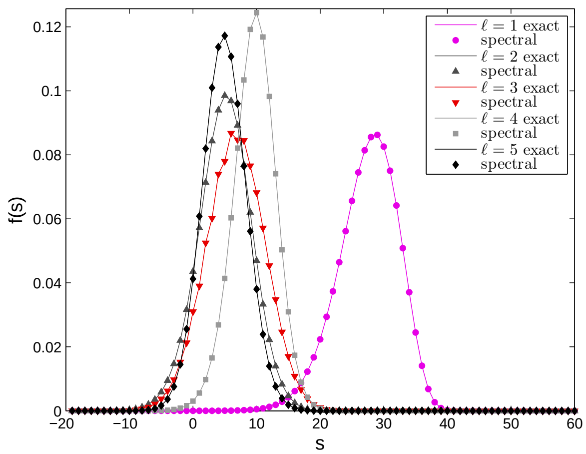

Figure 5.2: For different delays ℓ, densities computed at time $n = 60$ using the spectral method (plotted using solid markers) agree to machine precision with densities computed using an enumerative exact method (plotted using lines of the same grayscales/colors as markers). Computed densities are for five different versions of the random walk (5.4), each with a different delay $\ell \in \{1, 2, 3, 4, 5\}$. The Bernoulli steps K_n were as follows: outcomes $\{1, -1\}$ with probabilities $\{0.3, 0.7\}$ when $Y_n \geq Y_{n-\ell}$, and outcomes $\{2, -2\}$ with probabilities $\{0.9, 0.1\}$ when $Y_n < Y_{n-\ell}$. Initial conditions were $Y_n = 0$ for $n \leq \ell$.

n (in solid squares) and different delays ℓ (in solid circles). All plots confirm the spectral method's accuracy up to machine precision. To obtain the pdf at $n = 80$ in Fig. 5.1, the modified tree method takes 1390.8 s and the spectral method takes 0.67 s. To obtain the pdf for $\ell = 1$ in Fig. 5.2, the tree method from (Bhat and Kumar, 2012) takes 0.09 s, the modified tree method takes 0.29 s, and the spectral method takes 0.09 s. All simulations were done using Matlab on an 8-core Intel i7 CPU. All codes used to produce the results in the Chapter are available for download [2]. In all the experiments reported in Fig. 5.1 and Fig. 5.2, the spectral method is the fastest. Note that all of these results use the initial conditions $Y_j \equiv 0$ for $0 \leq j \leq \ell$.

Next, we apply the spectral method to the SDDE

$$dY_t = \tanh(Y_t - Y_{t-3dt})dt + dW_t, \tag{5.11}$$

subject to deterministic initial conditions $\theta(t) = 0$ for $t \leq 3dt$. Approximating $\tanh(x)$ by $\sum_{r=1}^{R} \mu_r I_{[c_r, c_{r+1})}(x)$ and applying the weak Euler-Maruyama discretization, we get (5.3) with $\sigma_r = 1$ for all r, $\ell = 3$ and $Y_n = 0$ for $n \leq \ell$. The error in the cumulative distribution function (cdf) calculated using the spectral method depends on the parameters R, c_r, μ_r used to approximate the $\tanh$ function; for the results shown in Fig. 5.4, these parameters' exact values are given in our Matlab code [3]

Setting $\Delta t = 0.04$ and $\Delta s = 0.01$ for both the $R = 3$ and $R = 5$ approximations, we compare their accuracies in Fig. 5.4. In the top pane, we first plot in solid gray the empirical cdf obtained at $T = 2$ $(\Delta t = 0.04)$ by simulating $M = 10^8$ sample paths of the Euler-Maruyama discretization of (5.11)—this Monte Carlo (MC) run was performed purely to give a reference solution against which we compare the spectral method's solutions. In the same pane, we plot

[2] http://faculty.ucmerced.edu/hbhat/codes/ssdrw.tar.gz Refer to the README file for details.

[3] http://faculty.ucmerced.edu/hbhat/codes/ssdrw.tar.gz

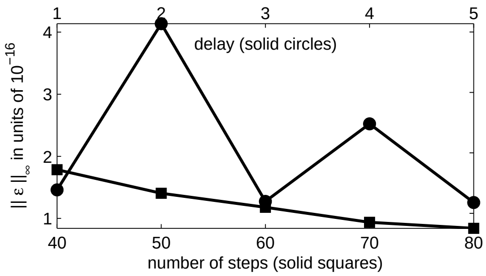

Figure 5.3: Infinity norm errors $\|f_{\mathrm{IFFT}}(n,s) - f(n,s)\|_\infty$ between the spectral and exact pdfs are at the level of machine precision, both as a function of time n and delay ℓ. Each solid square, one per value of time step n, corresponds to the $\|\cdot\|_\infty$ error between the solid markers (spectral) and lines (exact) in Fig. 5.1. Each solid circle, one per value of delay ℓ, corresponds to the $\|\cdot\|_\infty$ error between the solid markers (spectral) and lines (exact) in Fig. 5.2. All parameters are as in Fig. 5.1-5.2, respectively.

cdfs from spectral method simulations in dot-dashed gray ($R = 3$ approximation) and solid black ($R = 5$ approximation). In the bottom pane of Fig. 5.4, we plot the error between the MC cdf and the spectral method's cdfs in dot-dashed gray ($R = 3$) and solid black ($R = 5$). The maxima of the errors for the $R = 3$ and $R = 5$ approximations are, respectively, 0.0727 and 0.0337. The times taken to obtain the cdf through MC and the spectral method with $R = 3$ and $R = 5$ are 288.28 s, 0.56 s and 19.13 s, respectively. If we assume that the MC run is sufficiently fine-scale as to be close to the exact solution, these results suggest that the approximate solution will converge to the exact solution of the SDDE, as we increase R. We leave for future work a detailed discussion of convergence and optimal step function approximation.

In this Chapter, we have developed a spectral method to obtain the pdf of a delayed random walk that is both fast and accurate. As demonstrated, this method also shows promise to solve nonlinear SDDE. In future work, we plan to extend the spectral method to solve second-order and/or oscillatory SDDE (Kim et al., 1999; Barrio et al., 2006).

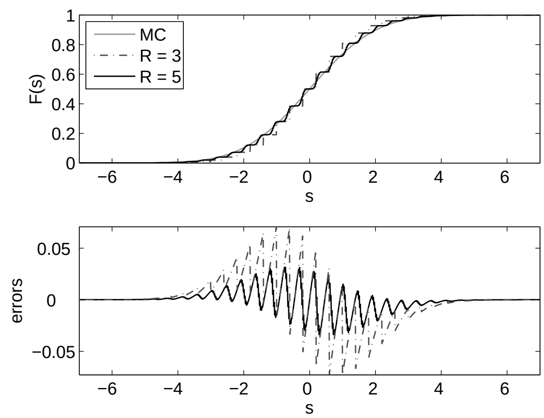

Figure 5.4: For the nonlinear SDDE (5.11), the accuracy of the cdf computed via the spectral method increases as we increase the number of piecewise constant branches R used to approximate the $\tanh$ function. In the top pane, we plot in solid gray a fine-scale reference cdf obtained through Monte Carlo (MC) simulation with $\Delta t = 0.04$ and 10^8 sample paths. In dot-dashed gray and solid black, we plot the cdfs obtained using the spectral method with $R = 3$ and $R = 5$ approximations, respectively. In the bottom pane, we plot the pointwise errors between the spectral method cdfs and the MC cdf. The maximum error decreases from 0.0727 to 0.0337 as we go from $R = 3$ to $R = 5$. All plots are at time $t = 2$ for zero initial conditions.

Bibliography

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.

AitSahlia, F., Goswami, M., and Guha, S. (2010). American option pricing under stochastic volatility: an empirical evaluation. *Computational Management Science*, 7(2):189–206.

Alexander, C. (2004). Normal mixture diffusion with uncertain volatility: modelling short- and long-term smile effects. *Journal of Banking & Finance*, 28:2957–2980.

Alexander, C., Kaeck, A., and Nogueira, L. (2009). Model risk adjusted hedge ratios. *Journal of Futures Markets*, 29(11):1021–1049.

An, Y. and Suo, W. (2009). An empirical comparison of option-pricing models in hedging exotic options. *Financial Management*, 38:889–914.

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.

Bakshi, G., Cao, C., and Chen, Z. (1997). Empirical performance of alternative options pricing models. *Journal of Finance*, 52:2003–2049.

Bakshi, G., Cao, C., and Chen, Z. (2000). Pricing and hedging long–term options. *Journal of Econometrics*, 94(1-2):277–318.

Bakshi, G., Kapadia, N., and Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, 16:101–143.

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

Barrio, M., Burrage, k., Leier, A., and Tian, T. (2006). Oscillatory regulation of hes1: discrete stochastic delay modelling and simulation. *PLOS Compututational Biology*, 2(9):e117.

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

Bates, D. S. (1995). Testing option pricing models. NBER Working Papers 5129, National Bureau of Economic Research, Inc.

Bates, D. S. (2000). Post-’87 crash fears in the S&P 500 futures option market. *Journal of Econometrics*, 94(1-2):181–238.

Bates, D. S. (2003). Empirical option pricing: A retrospection. *Journal of Econometrics*, 116(1):387–404.

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.

Berrones, A. and Larralde, H. (2001). Simple model of a random walk with arbitrarily long memory. *Physical Review E*, 63(3):031109.

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

Bhat, H. S. and Kumar, N. (2012). Option pricing under a normal mixture distribution derived from the markov tree model. *European Journal of Operational Research*, 223(3):762–774.

Bhattacharya, R. N. and Waymire, E. C. (2009). *Stochastic processes with applications*. SAIM.

Birgin, E. G. and Martinez, J. M. (2008). Improving ultimate convergence of an augmented Lagrangian method. *Optimization Methods and Software*, 23(2):177–195.

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

Blair, B., Poon, S. H., and Taylor, S. J. (2002). Asymmetric and crash effects in stock volatility for the s&p 100 index and its constituents. *Applied Financial Economics*, 12(5):319–329.

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

Bratsun, D., Volfson, D., Tsimring, L. S., and Hasty, J. (2005). Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences*, 102(41):14593–14598.

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.

Briggs, W. L. and Henson, V. E. (1995). *The discrete Fourier transform*. siam.

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.

Brigo, D., Mercurio, F., and Sartorelli, G. (2003). Alternative asset-price dynamics and volatility smile. *Quantitative Finance*, 3:173–183.

Broadie, M. and Detemple, J. B. (2004). Option pricing: Valuation models and applications. *Management Science*, 50(9):1145–1177.

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.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

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, C. C., Chou, P. H., and Liao, T. H. (2012). Fitting and testing for the implied volatility curve using parametric models. *Journal of Futures Markets*, 32:1171–191.

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.

Chou, R. K., Chung, S. L., Hsiao, Y. J., and Wang, Y. H. (2011). The impact of liquidity on option prices. *Journal of Futures Markets*, 31:1116–1141.

Conn, A. R., Gould, N. I. M., and Toint, P. L. (1991). A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2):545–572.

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

Corrado, C. and Su, T. (1996). Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial research*, 19:175–192.

Cox, J. C., Ross, S. A., and Rubinstein, M. (1979). Option pricing—Simplified approach. *Journal of Financial Economics*, 7(3):229–263.

Csiszár, I. (2002). Large-scale typicality of Markov sample paths and consistency of MDL order estimators. *IEEE Transactions on Information Theory*, 48(6):1616–1628.

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

Deco, G., Jirsa, V., and McIntosh, A. R. (2010). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56.

Deco, G., Jirsa, V., McIntosh, A. R., Sporns, O., and Kötter, R. (2009). Key role of coupling, delay, and noise in resting brain fluctuations. *Proceedings of the National Academy of Sciences*, 106(25):10302–10307.

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*, 1:83–186.

Duan, J. C. and Simonato, J. G. (2001). American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics and Control*, 25:1689–1718.

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.

Fiorentini, G., Leon, A., and Rubio, G. (2002). Estimation and empirical performance of Heston's stochastic volatility model: the case of a thinly traded market. *Journal of Empirical Finance*, 9(2):225–255.

Florescu, I. and Viens, F. G. (2008). Stochastic volatility: option pricing using a multinomial recombining tree. *Applied Mathematical Finance*, 15(2):151–181.

Frank, T. D. (2005a). Delay fokker-planck equations, novikov's theorem, and boltzmann distributions as small delay approximations. *Physical Review E*, 72:011112.

Frank, T. D. (2005b). Delay Fokker-planck equations, perturbation theory, and data analysis for nonlinear stochastic systems with time delays. *Physical Review E*, 71(3):031106.

French, K. R. and Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics*, 17(1):5–26.

Friesen, G. C., Zhang, Y., and Zorn, T. S. (2012). Heterogeneous beliefs and risk-neutral skewness. *Journal of Financial and Quantitative Analysis*, 47:851–872.

Gabriel, K. R. and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel-Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88:90–95.

Galla, T. (2009). Intrinsic fluctuations in stochastic delay systems: theoretical description and application to a simple model of gene regulation. *Physical Review E*, 80(2):021909.

Garcia, R., Ghysels, E., and Renault, E. (2003). The econometrics of option pricing. *Available at SSRN 463860*.

García-Pelayo, R. (2007). Solution of the persistent, biased random walk. *Physica A: Statistical Mechanics and its Applications*, 384(2):143–149.

Gilli, M. and Schumann, E. (2010). Calibrating the Heston model with differential evolution. In *Applications of Evolutionary Computation*, volume 6025 of *Lecture Notes in Computer Science*, pages 242–250. Springer Berlin / Heidelberg.

Guillouzic, S., L'heureux, I., and Iongtin, A. (1999). Small delay approximation of stochastic delay differential equations. *Physical Review E*, 59(4):3970–3982.

Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *Wilmott magazine*, pages 84–108.

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.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *Siam Review*, 43(3):525–546.

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

Hull, J. C. and White, A. D. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42(2):281–300.

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.

Jama, M. (2011). *Monocular vision based localization and mapping*. PhD thesis, Kansas State University.

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.

Johnson, S. G. (2013). *The NLopt nonlinear-optimization package*.

Josic, K., Lopez, J. M., Ott, W., Shiau, L., and Bennett, M. R. (2011). Stochastic delay accelerates signaling in gene networks. *PLOS Computational Biology*, 7(11):e1002264.

Kaeck, A. (2012). Hedging surprises, jumps, and model misspecification: a risk management perspective on hedging S&P 500 options. *Review of Finance*.

Kamrad, B. and Ritchken, P. (1991). Multinomial approximating models for options with k state variables. *Management science*, 37(12):1640–1652.

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.

Kim, S., Park, S. H., and Pyo, H. B. (1999). Stochastic resonance in coupled oscillator systems with time delay. *Physical Review Letters*, 82:1620–1623.

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

Lamoureux, C. G. and Lastrapes, W. D. (1993). Forecasting stock-return variance: toward an understanding of stochastic implied volatilities. *Review of Financial Studies*, 6:293–326.

Leisen, D. P. J. (2000). Stock evolution under stochastic volatility. *The Journal of Derivatives*, 8(2):9–27.

LIFFE, N. (2006). Guidelines for determining settlement prices for financial option contracts.

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

Lo, A. W. and MacKinlay, A. C. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies*, 1(1):41–66.

Lo, A. W. and MacKinlay, A. C. (1990). When are contrarian profits due to stock market overreaction? *Review of Financial studies*, 3(2):175–205.

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.

Longtin, A. (2010). stochastic delay-differential equations. In atay, f. m., editor, *complex time-delay systems*, volume 16 of *understanding complex systems*, pages 177–195. springer berlin / heidelberg.

Luethi, D. and Breymann, W. (2013). *ghyp: A package on generalized hyperbolic distributions*. R package version 1.5.6.

MacKenzie, D. (2004). The big, bad wolf and the rational market: portfolio insurance, the 1987 crash and the performativity of economics. *Economy and society*, 33(3):303–334.

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.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton Series in Finance. Princeton University Press.

McQueen, G. and Thorley, S. (1991). Are stock returns predictable? a test using markov chains. *The Journal of Finance*, 46(1):239–263.

Mikhailov, S. and Nögel, U. (2003). Heston's stochastic volatility model-implementation, calibration and some extensions. *Wilmott magazine*.

Milton, J. G. (2011). The delayed and noisy nervous system: implications for neural control. *Journal of Neural Engineering*, 8(6):065005.

Mugler, A., Walczak, A. M., and Wiggins, C. H. (2009). Spectral solutions to stochastic models of gene expression with bursts and regulation. *Pysical Review E*, 80(4):041921.

Nandi, S. (1996). Pricing and hedging index options under stochastic volatility: an empirical examination. Working Paper 96-9, Federal Reserve Bank of Atlanta.

Nandi, S. (1998). How important is the correlation between returns and volatility in a stochastic volatility model? empirical evidence from pricing and hedging in the S&P 500 index options market. *Journal of Banking & Finance*, 22:589–610.

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.

Ohira, T. and Milton, J. G. (1995). Delayed random walks. *Physical Review E*, 52(3):3277–3280.

Ohira, T. and Milton, J. G. (2009). Delayed random walks: investigating the interplay between delay and noise. In Balachandran, B., Kalmár-nagy, T., and Gilsinn, D. E., editors, *Delay differential equations*, pages 305–335. Springer.

Ohira, T. and Yamane, T. (2000). Delayed stochastic systems. *Physical Review E*, 61(2):1247.

Primbs, J. A., Rathinam, M., and Yamada, Y. (2007). Option pricing with a pentanomial lattice model that incorporates skewness and kurtosis. *Applied Mathematical Finance*, 14(1):1–17.

R Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

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.

Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978. *Journal of Finance*, 40(2):455–480.

Rubinstein, M. (2012). Implied binomial trees. *The Journal of Finance*, 49(3):771–818.

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.

Smolen, P., Baxter, D. A., and Byrne, J. H. (2002). A reduced model clarifies the role of feedback loops and time delays in the *drosophila* circadian oscillator. *Biophysical Journal*, 83:2349–2359.

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

Tan, B. and Yılmaz, K. (2002). Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation. *European Journal of Operational Research*, 137(3):524–543.

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.

Tanaka, D. H., Yanagida, M., Zhu, Y., Mikami, S., Nagasawa, T., Miyazaki, J. I., Yanagawa, Y., Obata, K., and Murakami, F. (2009). Random walk behavior of migrating cortical interneurons in the marginal zone: time-lapse analysis in flat-mount cortex. *The Journal of Neuroscience*, 29(5):1300–1311.

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

Van der Straeten, E. and Naudts, J. (2006). A two-parameter random walk with approximate exponential probability distribution. *Journal of Physics A: Mathematical and General*, 39(23):7245.

Weiss, G. H. (2002). Some applications of persistent random walks and the telegrapher's equation. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):381–410.

West, G. (2005). Calibration of the SABR model in illiquid markets. *Applied Mathematical Finance*, 12(4):371–385.

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

Yamada, Y. and Primbs, J. A. (2002). Distributon-based option pricing on lattice asset dynamics models. *International Journal of Theoretical and Applied Finance*, 05(06):599–618.

Yan, S. (2011). Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics*, 99:216–233.

Ypma, J. (2011). *Introduction to nloptr: an R interface to NLopt*. R package version 0.8.9.

Zhao, B. and Hodges, S. D. (2012). Parametric modeling of implied smile functions: a generalized SVI model. *Review of Derivatives Research*, pages 1–25.

Appendix A

Maximum Likelihood Estimation in Markov Chains

A.1 MLE's for k-th order Markov chain with Q symbols

Assume that we have a sequence $\{X_1, \ldots, X_N\}$ generated by a k-th order Markov chain where every experiment has Q possible outcomes.

Assume that each X_j takes values from a set $\{A_m\}_{m=1}^{Q}$ of Q distinct symbols. For a k-th order Markov chain, X_j depends on k outcomes prior to the j-th outcome. As in the earlier case with two possible outcomes, here also the Markov property kicks in only if $j > k$. Let S_i be a subsequence of k outcomes prior to the j-th outcome. Since each of these k outcomes in the subsequence is drawn from the set $\{A_m\}_{m=1}^{Q}$, there are Q^k possible subsequences S_i. Let these sequences be denoted by $\{S_i\}_{i=1}^{Q^k}$. Let us scan the given sequence $\{X_j\}_{j=1}^{N}$ from left to right and record the following Q^{k+1} numbers:

$$\{\{n_{S_iA_m} = \text{\# of times we observe “}S_iA_m\text{”}\}_{m=1}^{Q}\}_{i=1}^{Q^k}$$

For $m = 1$ to $m = Q - 1$, let $p_{S_iA_m} = P(A_m|S_i)$. Let $p_{S_iA_Q} = P(A_Q|S_i) = 1 - \sum_{m=1}^{Q-1} p_{S_iA_m}$ and $p_k = P(\{X_j\}_{j=1}^{k})$. In words, p_k is the probability of observing the first k terms of the $\{X_j\}_{j=1}^{N}$ sequence. Putting everything together, the log likelihood for the whole sequence is

$$L = \log p_k + \sum_{i=1}^{Q^k} \left[\sum_{m=1}^{Q-1} n_{S_iA_m} \log(p_{S_iA_m}) \right] + \sum_{i=1}^{Q^k} n_{S_iA_Q} \log \left(1 - \sum_{m=1}^{Q-1} p_{S_iA_m} \right).$$

Let us maximize L over all $p_{S_iA_m}$. Taking partial derivatives of both sides with respect to $p_{S_iA_m}$ for one particular (m, i), we get

$$\frac{1}{L} \frac{\partial L}{\partial p_{S_iA_m}} = \frac{n_{S_iA_m}}{p_{S_iA_m}} - \frac{n_{S_iA_Q}}{1 - \sum_{m=1}^{Q-1} p_{S_iA_m}}.$$

Setting $\partial L / \partial p_{S_i A_m} = 0$ to maximize L, we get

$$\frac{n_{S_i A_m}}{p_{S_i A_m}} = \frac{n_{S_i A_Q}}{1 - \sum_{m=1}^{Q-1} p_{S_i A_m}}. \tag{A.1}$$

Note that for a particular value of i, the above equation represents a set of $Q-1$ linear equations in $Q-1$ unknowns (the probabilities to be estimated), which give us the following result:

$$\frac{n_{S_i A_1}}{p_{S_i A_1}} = \frac{n_{S_i A_2}}{p_{S_i A_2}} = ... = \frac{n_{S_i A_{Q-1}}}{p_{S_i A_{Q-1}}} = \frac{n_{S_i A_Q}}{1 - \sum_{m=1}^{Q-1} p_{S_i A_m}}. \tag{A.2}$$

Solving the linear system given by (A.1) and (A.2), we get the MLE for the transition probability:

$$\hat{p}_{S_i A_m} = \frac{n_{S_i A_m}}{\sum_{m=1}^{Q} n_{S_i A_m}}.$$

We can then use the collection of all $\hat{p}$'s to find the maximum value of the log likelihood L.

Appendix B

Expressions in the Markov Tree model

B.1 Expressions for $\mu_{1,2}$ and $\sigma_{1,2}$

All parameters required to compute the MT price given by equation (4.10) depend on three unobservable parameters σ, σ^+ and σ^- through $\mu_{1,2}$ and $\sigma_{1,2}$.

$$
\begin{aligned}
\mu_1 = {} & -l_0 + \frac{l_1(q_1-1)(q_{-1}+q_1) + l_{-1}(q_{-1}(q_{-1}-3q_1+3)+q_1-1)}{(q_{-1}-q_1+1)^2} \\
& + \frac{(n-1)(l_1 q_{-1}(1-2q_1) + l_{-1}(2q_{-1}-1)(q_1-1))}{q_{-1}-q_1+1}, \\
\sigma_1^2 = {} & -(n-1)q_{-1}(q_1-1)\Big[q_1\left(l_1^2(4q_{-1}(q_{-1}+2)-3) - 2l_1l_{-1}(4q_{-1}(q_{-1}+2)-9) + l_{-1}^2(4q_{-1}(q_{-1}+2)-11)\right) \\
& - 4(q_{-1}-1){q_1}^2(l_1-l_{-1})^2 - (q_{-1}-1)(l_1-3l_{-1})^2\Big](q_{-1}-q_1+1)^{-3} \\
& - (q_1-1)\Big[l_1^2\left({q_{-1}}^3 + {q_{-1}}^2(2q_1-1) + q_{-1}(q_1(9q_1-8)+1) + (q_1-1)q_1\right) \\
& - 2l_1l_{-1}\left({q_{-1}}^3 + {q_{-1}}^2(6q_1-5) + q_{-1}(5(q_1-2)q_1+3) - {q_1}^2 + q_1\right) \\
& + l_{-1}^2\left({q_{-1}}^3 + {q_{-1}}^2(10q_1-13) + q_{-1}((q_1-12)q_1+13) + (q_1-1)q_1\right)\Big](q_{-1}-q_1+1)^{-4} \\
\mu_2 = {} & l_0 + \frac{l_1\left(q_{-1}(3q_1-2) - (q_1-1)^2\right) - l_{-1}q_{-1}(q_{-1}+q_1-2)}{(q_{-1}-q_1+1)^2} \\
& + \frac{(n-1)(l_1 q_{-1}(1-2q_1) + l_{-1}(2q_{-1}-1)(q_1-1))}{q_{-1}-q_1+1}, \\
\sigma_2^2 = {} & -(n-1)q_{-1}(q_1-1)\Big[q_1\left(l_1^2(4q_{-1}(q_{-1}+2)-3) - 2l_1l_{-1}(4q_{-1}(q_{-1}+2)-9) + l_{-1}^2(4q_{-1}(q_{-1}+2)-11)\right) \\
& - 4(q_{-1}-1){q_1}^2(l_1-l_{-1})^2 - (q_{-1}-1)(l_1-3l_{-1})^2\Big](q_{-1}-q_1+1)^{-3} \\
& - q_{-1}\Big[l_1^2\left({q_{-1}}^2(q_1-2) + q_{-1}(10(q_1-1)q_1+1) + {q_1}^3 - q_1\right) \\
& - 2l_1l_{-1}\left({q_{-1}}^2(5q_1-4) + q_{-1}(6(q_1-2)q_1+5) + (q_1-3)(q_1-1)q_1\right) \\
& + l_{-1}^2\left({q_{-1}}^2(9q_1-10) + q_{-1}(2(q_1-7)q_1+13) + (q_1-1)((q_1-3)q_1+4)\right)\Big](q_{-1}-q_1+1)^{-4}
\end{aligned}
$$

where

$$l_1 = \sigma^+\sqrt{\Delta t}, \quad l_{-1} = \sigma^-\sqrt{\Delta t}, \quad l_0 = \sigma\sqrt{\Delta t}$$
$$q_k = \frac{e^{r\Delta t} - e^{-l_k}}{e^{l_k} - e^{-l_k}}, \quad \text{for} \quad k \in \{-1, 0, 1\}$$

and $\Delta t = T/N$ where T is the time in years to expiry and N is the number of steps in the MT.

B.2 Gradient vector for the MT objective function.

For an option with strike K and time to expiration T, the error between the MT price and the market price is

$$\epsilon = V_\Theta^{K,T} - F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})$$

where $V_\Theta^{K,T}$ is the market price of the option, $F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})$ is the MT model price, and $\boldsymbol{\beta}^{\mathrm{MT}} = (\sigma, \sigma^+, \sigma^-)$. From the above equation, we get

$$\frac{\partial \epsilon}{\partial \sigma} = -\frac{\partial F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})}{\partial \sigma}, \quad \frac{\partial \epsilon}{\partial \sigma^\pm} = -\frac{\partial F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})}{\partial \sigma^\pm}.$$

Using (4.10), we express $F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})$ as

$$F^{\mathrm{MT}}(\mathbf{x}, \boldsymbol{\beta}^{\mathrm{MT}})e^{rT} = q_0 f_1(\mu_1, \sigma_1) + (1 - q_0) f_2(\mu_2, \sigma_2), \tag{B.1}$$

where

$$f_i(\mu_i, \sigma_i) = S_0 \exp\left(\frac{\sigma_i^2}{2} + \mu_i\right)\Phi(x_i) - K\Phi(x_{i+2}),$$

and $x_1, \ldots, x_4$ are given in (4.11) and μ_1, μ_2, σ_1, σ_2 are given in Appendix B.1.

Partial with respect to first parameter $\boldsymbol{\partial \mathrm{F}^{\mathrm{MT}}/\partial \sigma}$. To calculate $\partial \epsilon_m/\partial \sigma$, we first need to calculate $\partial F^{\mathrm{MT}}/\partial \sigma$:

$$\begin{aligned}
e^{rT}\frac{\partial F^{\mathrm{MT}}}{\partial \sigma} &= q_0\left(\frac{\partial f_1}{\partial \sigma} - \frac{\partial f_2}{\partial \sigma}\right) + \frac{\partial q_0}{\partial \sigma}(f_1 - f_2) + \frac{\partial f_2}{\partial \sigma} \\
\frac{\partial q_0}{\partial \sigma} &= -\sqrt{\Delta t}e^{\sigma\sqrt{\Delta t}}\left(\frac{e^{r\Delta t} - 2e^{\sigma\sqrt{\Delta t}} + e^{r\Delta t + 2\sigma\sqrt{\Delta t}}}{(e^{2\sigma\sqrt{\Delta t}} - 1)^2}\right) \\
\frac{\partial f_i}{\partial \sigma} &= S_0 \exp\left(\frac{\sigma_i^2}{2} + \mu_i\right)\left(N(x_i)\frac{\partial x_i}{\partial \sigma} + \left(\frac{1}{2}\frac{\partial \sigma_i^2}{\partial \sigma} + \frac{\partial \mu_i}{\partial \sigma}\right)\Phi(x_i)\right) - KN(x_{i+2})\frac{\partial x_{i+2}}{\partial \sigma}.
\end{aligned}$$

From the definition of x_i's in section 4.2.3, we get

$$\frac{\partial x_1}{\partial \sigma} = \frac{\partial x_3}{\partial \sigma} = \frac{1}{\sigma_1}\frac{\partial \mu_1}{\partial \sigma}, \text{ and } \quad \frac{\partial x_2}{\partial \sigma} = \frac{\partial x_4}{\partial \sigma} = \frac{1}{\sigma_2}\frac{\partial \mu_2}{\partial \sigma},$$

and from the expressions for $\mu_{1,2}$ and $\sigma^2_{1,2}$, we know that

$$\frac{\partial \sigma^2_{1,2}}{\partial \sigma} = 0, \quad \frac{\partial \mu_1}{\partial \sigma} = \sqrt{dt}, \quad \frac{\partial \mu_2}{\partial \sigma} = -\sqrt{dt}.$$

Partial with respect to second and third parameters $\boldsymbol{\partial \mathrm{F}^{\mathrm{MT}}/\partial \sigma^{\pm}}$. We now move on to calculating $\partial \epsilon_m/\partial \sigma^{\pm}$ for which we first need to calculate $\partial F^{\mathrm{MT}}/\partial \sigma^{\pm}$. We get

$$e^{rT}\frac{\partial F^{\mathrm{MT}}}{\partial \sigma^{\pm}} = q_0\left(\frac{\partial f_1}{\partial \sigma^{\pm}} - \frac{\partial f_2}{\partial \sigma^{\pm}}\right) + \frac{\partial q_0}{\partial \sigma^{\pm}}(f_1 - f_2) + \frac{\partial f_2}{\partial \sigma^{\pm}},$$

where $\partial q_0/\partial \sigma^{\pm} = 0$. The above expressions consists of the terms $\partial f_1/\partial \sigma^{\pm}$ and $\partial f_2/\partial \sigma^{\pm}$ that can in turn be expressed as

$$\frac{\partial f_i}{\partial \sigma^{\pm}} = S_0 \exp\left(\frac{\sigma_i^2}{2} + \mu_i\right)\left(N(x_i)\frac{\partial x_i}{\partial \sigma^{\pm}} + \left(\frac{1}{2}\frac{\partial \sigma_i^2}{\partial \sigma^{\pm}} + \frac{\partial \mu_i}{\partial \sigma^{\pm}}\right)\Phi(x_i)\right) - KN(x_{i+2})\frac{\partial x_{i+2}}{\partial \sigma^{\pm}}.$$

For $i = \{1, 2\}$, $\partial x_i/\partial \sigma^{\pm}$ is

$$\frac{\partial x_i}{\partial \sigma^{\pm}} = \frac{\partial x_{i+2}}{\partial \sigma^{\pm}} + \frac{1}{2\sigma_i}\frac{\partial \sigma_i^2}{\partial \sigma^{\pm}}$$

and $\partial x_{3,4}/\partial \sigma^{\pm}$ are

$$\frac{\partial x_{i+2}}{\partial \sigma^{\pm}} = -\frac{1}{2\sigma_i^3}\mu_i\frac{\partial \sigma_i^2}{\partial \sigma^{\pm}} + \frac{1}{\sigma_i}\frac{\partial \mu_i}{\partial \sigma^{\pm}} - \frac{1}{2\sigma_i^3}\log\left(\frac{S_0}{X}\right),$$

where X is the strike price of the option. To evaluate the above expression, we need the partials of $\mu_{1,2}$ and $\sigma^2_{1,2}$ with respect to $\sigma^{\pm}$.

$$\frac{\partial \mu_{1,2}}{\partial \sigma^{+}} = \frac{\partial \mu_{1,2}}{\partial l_1}\frac{\partial l_1}{\partial \sigma^{+}} + \frac{\partial \mu_{1,2}}{\partial q_1}\frac{\partial q_1}{\partial \sigma^{+}}, \quad \frac{\partial \mu_{1,2}}{\partial \sigma^{-}} = \frac{\partial \mu_{1,2}}{\partial l_{-1}}\frac{\partial l_{-1}}{\partial \sigma^{-}} + \frac{\partial \mu_{1,2}}{\partial q_{-1}}\frac{\partial q_{-1}}{\partial \sigma^{-}},$$
$$\frac{\partial \sigma^2_{1,2}}{\partial \sigma^{+}} = \frac{\partial \sigma^2_{1,2}}{\partial l_1}\frac{\partial l_1}{\partial \sigma^{+}} + \frac{\partial \sigma^2_{1,2}}{\partial q_1}\frac{\partial q_1}{\partial \sigma^{+}}, \quad \frac{\partial \sigma^2_{1,2}}{\partial \sigma^{-}} = \frac{\partial \sigma^2_{1,2}}{\partial l_{-1}}\frac{\partial l_{-1}}{\partial \sigma^{-}} + \frac{\partial \sigma^2_{1,2}}{\partial q_{-1}}\frac{\partial q_{-1}}{\partial \sigma^{-}}.$$

The above expressions then depend on

$$\frac{\partial \mu_1}{\partial l_1} = \frac{(2q_1 - 1)q_{-1}(n-1)}{1 + q_{-1} - q_1} + \frac{(1 - q_1)(q_1 + q_{-1})}{(1 + q_{-1} - q_1)^2},$$
$$\frac{\partial \mu_2}{\partial l_1} = \frac{(2q_1 - 1)q_{-1}(n-1)}{1 + q_{-1} - q_1} + \frac{(1 - q_1)^2 - q_{-1}(3q_1 - 2)}{(1 + q_{-1} - q_1)^2},$$
$$\frac{\partial \mu_1}{\partial l_{-1}} = \frac{(2q_{-1} - 1)(1 - q_1)(n-1)}{1 + q_{-1} - q_1} - \frac{q_1 + q_{-1}(3 + q_{-1} - 3q_1) - 1}{(1 + q_{-1} - q_1)^2},$$
$$\frac{\partial \mu_2}{\partial l_{-1}} = \frac{(2q_{-1} - 1)(1 - q_1)(n-1)}{1 + q_{-1} - q_1} - \frac{q_{-1}(2 - q_1 - q_{-1})}{(1 + q_{-1} - q_1)^2}.$$

We also know that $\partial l_1/\partial\sigma^+ = \sqrt{\Delta t}$ and $\partial l_{-1}/\partial\sigma^- = \sqrt{\Delta t}$. To completely specify the partials of $\mu_{1,2}$ with respect to $l_{-1,1}$ we only need the partials of the tree probabilities $q_{-1,1}$ with respect to $\sigma^{\pm}$. We calculate:

$$\frac{\partial q_{-1,1}}{\partial\sigma^{\pm}} = -\sqrt{\Delta t}e^{\sigma^{\pm}\sqrt{\Delta t}}\left(\frac{e^{r\Delta t} - 2e^{\sigma^{\pm}\sqrt{\Delta t}} + e^{r\Delta t+2\sigma^{\pm}\sqrt{\Delta t}}}{(e^{2\sigma^{\pm}\sqrt{\Delta t}}-1)^2}\right).$$

We are now left with the task of defining the partials of $\sigma^2_{1,2}$ with respect to $l_{1,2}$.

$$\begin{aligned}\frac{\partial\sigma_1^2}{\partial l_1} = {} & \frac{(n-1)q_{-1}(1-q_1)\left((-2l_{-1}(-9+4q_{-1}(2+q_{-1})) + 2l_1(-3+4q_{-1}(2+q_{-1})))q_1\right)}{(q_{-1}-q_1+1)^3} \\ & -\frac{(n-1)q_{-1}(1-q_1)\big(2(l_1-3l_{-1})(-1+q_{-1}) + 8(l_1-l_{-1})(-1+q_{-1})(q_1)^2\big)}{(q_{-1}-q_1+1)^3} \\ & +\frac{(1-q_1)\left(-2l_{-1}((q_{-1})^3 + q_1 - (q_1)^2 + (q_{-1})^2(-5+6q_1) + q_{-1}(3+5(-2+q_1)q_1))\right)}{(q_{-1}-q_1+1)^4} \\ & +\frac{(1-q_1)\left(2l_1((q_{-1})^3 + (-1+q_1)q_1 + (q_{-1})^2(-1+2q_1) + q_{-1}(1+q_1(-8+9q_1)))\right)}{(q_{-1}-q_1+1)^4}.\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial\sigma_1^2}{\partial l_{-1}} = {} & \frac{(n-1)q_{-1}(1-q_1)\left((2l_{-1}(-11+4q_{-1}(2+q_{-1})) - 2l_1(-9+4q_{-1}(2+q_{-1})))q_1\right)}{(q_{-1}-q_1+1)^3} \\ & +\frac{(n-1)q_{-1}(1-q_1)\left(6(l_1-3l_{-1})(-1+q_{-1}) + 8(l_1-l_{-1})(-1+q_{-1})(q_1)^2\right)}{(q_{-1}-q_1+1)^3} \\ & -\frac{(1-q_1)\left(2l_1((q_{-1})^3 + q_1 - (q_1)^2 + (q_{-1})^2(-5+6q_1) + q_{-1}(3+5(-2+q_1)q_1))\right)}{(q_{-1}-q_1+1)^4} \\ & +\frac{(1-q_1)\left(2l_{-1}((q_{-1})^3 + (-1+q_1)q_1 + (q_{-1})^2(-13+10q_1) + q_{-1}(13+(-12+q_1)q_1)\right)}{(q_{-1}-q_1+1)^4}\end{aligned}$$

Finally, we can express $\partial\sigma_2^2/\partial l_{1,2}$ in terms of $\partial\sigma_1^2/\partial l_{1,2}$ as

$$\begin{aligned}\frac{\partial\sigma_2^2}{\partial l_1} = \frac{\partial\sigma_1^2}{\partial l_1} & + \frac{2l_1((q_{-1})^2 - (-1+q_1)q_1 + q_{-1}(-1+8q_1-8(q_1)^2))}{(1+q_{-1}-q_1)^3} \\ & +\frac{2l_{-1}(-(-1+q_1)q_1 + (q_{-1})^2(-3+4q_1) + q_{-1}(3-8q_1+4(q_1)^2))}{(1+q_{-1}-q_1)^3}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\sigma_2^2}{\partial l_{-1}} = \frac{\partial\sigma_1^2}{\partial l_{-1}} & + \frac{2l_1(-(-1+q_1)q_1 + (q_{-1})^2(-3+4q_1) + q_{-1}(3-8q_1+4(q_1)^2))}{(1+q_{-1}-q_1)^3} \\ & +\frac{2l_{-1}((q_{-1})^2(9-8q_1) - (-1+q_1)q_1 + q_{-1}(-9+8q_1))}{(1+q_{-1}-q_1)^3}.\end{aligned}$$

We now move on to the expressions for the partials of the variances with respect to the tree probabilities:

$$\frac{\partial\sigma_1^2}{\partial q_1} = (1+q_{-1}-q_1)^{-5}\Big[2l_1l_{-1}(-(-1+q_1)^2(1+q_1)+4(-1+n)(q_{-1})^5(-1+2q_1)$$
$$+q_{-1}(-1+q_1)(3-15q_1+4(q_1)^2+n(-3+2q_1+(q_1)^2))$$
$$+(q_{-1})^4(n(-19+36q_1-16(q_1)^2)+4(5-9q_1+4(q_1)^2))$$
$$+(q_{-1})^2(9-15q_1-8(q_1)^2+12(q_1)^3+n(13-36q_1+35(q_1)^2-12(q_1)^3))$$
$$+(q_{-1})^3(-9+20(q_1)^2-8(q_1)^3+n(-5+15q_1-20(q_1)^2+8(q_1)^3)))$$
$$+l_{-1}^2(-(-1+q_1)^2(1+q_1)-4(-1+n)(q_{-1})^5(-1+2q_1)$$
$$-q_{-1}(-1+q_1)(19-9q_1-2(q_1)^2+n(7-10q_1+3(q_1)^2))$$
$$+(q_{-1})^4(-26+36q_1-16(q_1)^2+n(25-36q_1+16(q_1)^2))-(q_{-1})^3(-35+18q_1+20(q_1)^2-8(q_1)^3$$
$$+n(9+5q_1-20(q_1)^2+8(q_1)^3))+(q_{-1})^2(-31+15q_1+26(q_1)^2-12(q_1)^3$$
$$+n(-23+60q_1-49(q_1)^2+12(q_1)^3)))+l_1^2(-(-1+q_1)^2(1+q_1)-4(-1+n)(q_{-1})^5(-1+2q_1)$$
$$+q_{-1}(-1+q_1)(5+n+q_1-6nq_1-14(q_1)^2+5n(q_1)^2)$$
$$+(q_{-1})^4(-2(9-18q_1+8(q_1)^2)+n(17-36q_1+16(q_1)^2))$$
$$+(q_{-1})^3(-9+22q_1-20(q_1)^2+8(q_1)^3+n(15-29q_1+20(q_1)^2-8(q_1)^3))$$
$$+(q_{-1})^2(-11+27q_1-6(q_1)^2-12(q_1)^3+n(1+12q_1-25(q_1)^2+12(q_1)^3))\Big]$$

$$\frac{\partial\sigma_1^2}{\partial q_{-1}} = (1+q_{-1}-q_1)^{-5}(q_1-1)\Big[-(1+q_{-1}-q_1)(-3(-1+n)q_{-1}(-(l_1-3l_{-1})^2(-1+q_{-1})$$
$$+(l_{-1}^2(-11+4q_{-1}(2+q_{-1}))-2l_1l_{-1}(-9+4q_{-1}(2+q_{-1}))+l_1^2(-3+4q_{-1}(2+q_{-1})))q_1$$
$$-4(l_1-l_{-1})^2(-1+q_{-1})(q_1)^2)+l_1^2(1+3(q_{-1})^2-8q_1+9(q_1)^2+q_{-1}(-2+4q_1))$$
$$-2l_1l_{-1}(3+3(q_{-1})^2+5(-2+q_1)q_1+2q_{-1}(-5+6q_1))+l_{-1}^2(13+3(q_{-1})^2-12q_1+(q_1)^2$$
$$+q_{-1}(-26+20q_1)))+(-1+n)(1+q_{-1}-q_1)^2(l_{-1}^2(-1+2q_{-1})(9-(11+6q_{-1})q_1+4(q_1)^2)$$
$$+l_1^2(-1+3q_1-12(q_{-1})^2q_1-4(q_1)^2+2q_{-1}(1-8q_1+4(q_1)^2))$$
$$+2l_1l_{-1}(3-9q_1+12(q_{-1})^2q_1+4(q_1)^2-2q_{-1}(3-8q_1+4(q_1)^2)))$$
$$+4(l_{-1}^2((q_{-1})^3+(-1+q_1)q_1+(q_{-1})^2(-13+10q_1)+q_{-1}(13+(-12+q_1)q_1))$$
$$-2l_1l_{-1}((q_{-1})^3+q_1-(q_1)^2+(q_{-1})^2(-5+6q_1)+q_{-1}(3+5(-2+q_1)q_1))$$
$$+l_1^2((q_{-1})^3+(-1+q_1)q_1+(q_{-1})^2(-1+2q_1)+q_{-1}(1+q_1(-8+9q_1))))\Big]$$

$$\frac{\partial\sigma_2^2}{\partial q_1} = \frac{\partial\sigma_1^2}{\partial q_1}-(1+q_{-1}-q_1)^{-4}\Big[l_{-1}^2(-1+8(q_{-1})^3+(q_1)^2-2q_{-1}(-9+7q_1)+(q_{-1})^2(-27+16q_1))$$
$$-2l_1l_{-1}(1+4(q_{-1})^3-(q_1)^2+(q_{-1})^2(-13+16q_1)+2q_{-1}(1-5q_1+2(q_1)^2))$$
$$+l_1^2(-1+(q_1)^2+(q_{-1})^2(-11+16q_1)+2q_{-1}(-3+q_1+4(q_1)^2))\Big]$$

$$\frac{\partial \sigma_2^2}{\partial q_{-1}} = \frac{\partial \sigma_1^2}{\partial q_{-1}} + (1 + q_{-1} - q_1)^{-4} \Big[l_{-1}^2(-9 + 14q_1 - 5(q_1)^2 + (q_{-1})^2(-9 + 8q_1) + 2q_{-1}(18 - 25q_1 + 8(q_1)^2))$$
$$-2l_1 l_{-1}(-3 + 14q_1 - 15(q_1)^2 + 4(q_1)^3 + (q_{-1})^2(-3 + 4q_1) + 2q_{-1}(6 - 15q_1 + 8(q_1)^2))$$
$$-l_1^2(1 + (q_{-1})^2 - 6q_1 + 13(q_1)^2 - 8(q_1)^3 - 2q_{-1}(2 - 9q_1 + 8(q_1)^2))\Big]$$

B.3 Delta neutral in the MT model.

The option Delta for the MT model is given by

$$\frac{\partial F^{\mathrm{MT}}}{\partial S_0} = e^{-rT}\left(q_0 \frac{\partial f_1}{\partial S_0} + (1 - q_0)\frac{\partial f_2}{\partial S_0}\right),$$

where

$$\frac{\partial f_i}{\partial S_0} = \Phi(x_i) \exp\left(\frac{\sigma_i^2}{2} + \mu_i\right) + \frac{1}{\sigma_i} N(x_i) \exp\left(\frac{\sigma_i^2}{2} + \mu_i\right) - N(x_{i+2}) \frac{K}{S_0 \sigma_i} \tag{B.2}$$