

ISSN 1440-771X



**DEPARTMENT OF ECONOMETRICS  
AND BUSINESS STATISTICS**

**Implicit Bayesian Inference Using Option Prices**

**Gael M Martin, Catherine S Forbes and Vance L Martin**

**Working Paper 5/2003**

# Implicit Bayesian Inference Using Option Prices

Gael M. Martin<sup>a</sup>, Catherine S. Forbes<sup>a</sup> and Vance L. Martin<sup>b\*</sup>

a. Department of Econometrics and Business Statistics, PO Box 11E  
Monash University, Vic., 3800, Australia; Email: Gael.Martin@BusEco.monash.edu.au.

Phone: 61 3 9905 1189; Fax: 61 3 9905 5474.

b. Department of Economics, University of Melbourne.

First Draft, July, 2000  
(Monash University Working Paper 5/2000)  
This Version, February, 2003

## Abstract

A Bayesian approach to option pricing is presented, in which posterior inference about the underlying returns process is conducted implicitly via observed option prices. A range of models allowing for conditional leptokurtosis, skewness and time-varying volatility in returns are considered, with posterior parameter distributions and model probabilities backed out from the option prices. Models are ranked according to several criteria, including out-of-sample fit, predictive and hedging performance. The methodology accommodates heteroscedasticity and autocorrelation in the option pricing errors, as well as regime shifts across contract groups. The method is applied to intraday option price data on the S&P500 stock index for 1995. Whilst the results provide support for models which accommodate leptokurtosis, no one model dominates according to all criteria considered.

Keywords: Bayesian Option Pricing; Leptokurtosis; Skewness; GARCH Option Pricing; Option Price Prediction; Hedging Errors.

JEL Classifications: C11, C16, G13.

---

\*This research has been supported by an Australian Research Council Large Grant. The authors would like to thank an Associate Editor and two referees for very constructive and insightful comments on an earlier draft of the paper. All remaining errors are the responsibility of the authors.

# 1 Introduction

An option is a contingent claim whose theoretical price is dependent upon the process assumed for returns on the underlying asset on which the option is written. Observed market option prices thus contain information on this process which is potentially different from and more complete than, information contained in an historical time series on returns; see, for example, Pastorello, Renault and Touzi (2000). In this paper, a methodology is presented for conducting implicit inference about a range of models for the underlying returns process, using option price data. The methodology is based on the Bayesian paradigm and involves the production of both posterior densities for the parameters of the alternative models and posterior model probabilities. The models considered allow for both time-varying conditional volatility, using the Generalized Autoregressive Conditional Heteroscedasticity (*GARCH*) framework of Engle (1982) and Bollerslev (1986), and leptokurtosis and skewness in the conditional distribution of returns, using the frameworks of Lye and Martin (1993, 1994) and Fernandez and Steel (1998). The generalized local risk-neutral valuation method of Duan (1999) is used as the basis for defining the pertinent risk-neutral process in the estimation of all models which assume a nonnormal conditional distribution. An important feature of the proposed framework is that it nests the option pricing model of Black and Scholes (1973), in which returns are assumed to be normally distributed with constant volatility.

To assess the out-of-sample performance of the different parametric models, fit and predictive densities are produced. The hedging performance of the different models is also gauged via the construction of posterior densities for the hedging errors. The posterior densities for the model parameters and the posterior model probabilities are based on the prices of option contracts on the S&P500 stock index recorded during the first 239 trading days of 1995. The out-of-sample fit, predictive and hedging error assessments are based on data recorded during the week immediately succeeding the end of the estimation period.<sup>1</sup>

Most of the existing statistical work on option prices is based on either the classical paradigm or on a simple application of statistical fit. Engle and Mustafa (1992), Sabbatini and Linton (1998) and Heston and Nandi (2000) minimize the sum of squared deviations between observed and theoretical option prices to estimate the parameters of *GARCH* processes. Dumas, Fleming and Whaley (1998) adopt a similar approach using deterministic volatility models, whilst Jackwerth and Rubenstein (2001) use measures of fit to infer a variety of deterministic and stochastic volatility models. Bates (2000), Chernov and Ghysels (2000),

---

<sup>1</sup>The data has been obtained from the Berkeley Options Database.

and Pan (2002) use more formal classical methods to produce implicit estimates of the parameters of stochastic volatility models, based on the assumption of conditional normality for the returns process. In Lim et al (1998), Bollerslev and Mikkelsen (1999), Duan (1999) and Hafner and Herwartz (2001), *GARCH* models are augmented with nonnormal conditional errors and the implications of such models for option pricing investigated, again within a classical inferential framework. In Corrado and Su (1997), Dutta and Babbel (2002) and Lim, Martin and Martin (2002a), option prices are used to conduct classical implicit estimation of returns models which accommodate skewness and leptokurtosis, with a time-varying volatility component also specified in the case of Lim, Martin and Martin (2002a). Significant option-implied skewness and excess kurtosis is found in all cases, with the link between these features and implied volatility smiles highlighted in Lim, Martin and Martin (2002a). Backus, Foresis, Li and Wu (1997) also focus on the connection between volatility smiles and departures from lognormality in the underlying spot price process. Lim, Martin and Martin (2002b) extend this type of modelling approach to the less usual case of volatility frowns, linking this feature to the presence of thin-tailed underlying returns processes.

Some Bayesian analyses have been performed. Boyle and Ananthanarayanan (1977) and Korolyi (1993) conduct Bayesian inference in an option pricing framework using returns data, with attention restricted to the Black-Scholes (*BS*) model. Bauwens and Lubrano (2002) also use returns data to conduct Bayesian inference, but allow for deviations from the *BS* assumptions. In line with the present paper, Jacquier and Jarrow (2000) conduct Bayesian inference using observed option prices. Unlike our approach, however, in which the option price data is used to estimate and rank a full set of parametric returns models, Jacquier and Jarrow focus on the *BS* model, catering for the misspecification of that model nonparametrically. We also use a richer specification for the option pricing errors than do the latter authors. Jones (2000), Eraker (2001), Forbes, Martin and Wright (2002) and Polson and Stroud (2002) use option prices to estimate stochastic volatility models for returns, applying Bayesian inferential methods. In all cases, however, the assumption of conditional normality is maintained.

The paper is organized as follows. Section 2 discusses the application of the Bayesian statistical paradigm to option pricing. Alternative option price models that allow for time-varying volatility and nonnormality in the conditional distribution of returns are formulated in Section 3, along with the appropriate risk-neutral adjustments. In Section 4, implicit Bayesian inference based on option price data on the S&P500 index is illustrated. Posterior quantities are reported, together with summary measures of the fit, predictive and hedging

distributions for the different models. The empirical results provide evidence which favours a fat-tailed model, with both point and interval estimates indicating that the option prices have factored in the assumption of a returns distribution with excess kurtosis. The model which allows for excess kurtosis has the largest posterior probability and the best out-of-sample performance according to most criteria considered. There is evidence of a small amount of negative skewness being factored into the option prices, more than would be warranted by consideration of the skewness properties of returns on the index during the relevant time period. However, little posterior weight is assigned to the model which departs from normality only in the sense of being skewed. The *GARCH* models are also assigned little posterior weight in comparison with the constant volatility models, although within the *GARCH* class there is a clear hierarchy, with the models which allow for conditional nonnormality performing better overall than the model which adopts a normal conditional distribution for returns. The hedging results suggest that the hedging errors for all models are insubstantial. Some conclusions are drawn in Section 5.

## 2 Bayesian Inference in an Option Pricing Framework

The price of an option written on a non-dividend paying asset is the expected value of the discounted payoff of the option. For a European call option, the price is

$$q = E_t \left[ e^{-r\tau} \max(S_T - K, 0) \right], \quad (1)$$

where  $E_t$  is the conditional expectation, based on information at time  $t = T - \tau$ , taken with respect to the risk-neutral probability measure; see Hull (2000). The notation used in (1) is defined as follows:

- $T$  = the time at which the option is to be exercised;
- $\tau$  = the length of the option contract;
- $K$  = the exercise price;
- $S_T$  = the spot price of the underlying asset at the time of maturity;
- $r$  = the risk-free interest rate assumed to hold over the life of the option.

The option price is thus a function of certain observable quantities, namely  $r$ ,  $K$  and  $\tau$ . As the expectation is evaluated at time  $t$ , it is also a function of the observable level of the spot price prevailing at that time,  $S_t$ . Since the option price involves the evaluation of the expected payoff at the time of maturity, the price depends on (i) the assumed stochastic

process for  $S_t$ , or alternatively, on the assumed distribution for returns on the asset; and (ii) the values assigned to the unknown parameters of that underlying process. In this paper, we explicitly allow for the uncertainty associated with both (i) and (ii), by producing respectively posterior probabilities for a range of alternative models and posterior probability distributions for the model specific parameters.

Posterior inferences are to be produced implicitly from observed market option prices. For this to occur, option prices need to be assigned a particular distributional model. In this paper, a very general model is adopted, whereby option pricing errors are allowed to be serially correlated across days and heterogeneous across both time and moneyness category. As the empirical application focusses only on short-term options, with less than a month and a half to expiry, no allowance is made for variation across maturity category.

Let  $C_{ijt}$  denote the price of option contract  $i$  in moneyness category  $j$ , observed at time  $t$ , where moneyness group  $j$ ,  $j = 1, 2, \dots, J$ , is defined according to

$$m_j < \frac{S_t}{K_{ij}} < m_{j+1},$$

with  $K_{ij}$  denoting the exercise price associated with  $C_{ijt}$ . The number of groups and the location of segment boundaries,  $m_j$ ,  $j = 1, 2, \dots, J$ , are chosen to accord with the main moneyness groups in the data. More details of this are provided in Section 4. Although synchronous recording of the spot and option prices is a feature of the empirical data, we do not attempt to model movements in the underlying spot price process across the day. Rather, we produce inferences, via observed option prices, on the day-to-day movements in  $S_t$ , or, in other words, inference on the daily returns process. Hence, we attempt to minimize the within-day variation in  $S_t$  in the option price sample by selecting a cross section of option prices observed at (approximately) the same time on each day,  $t$ , where  $t = 1, 2, \dots, n$ , and  $n$  is the number of trading days used in the estimation sample.<sup>2</sup> The number of observations in each moneyness group at each point in time,  $n_{jt}$ , varies. Letting  $i = 1, 2, \dots, n_{jt}$ ,  $j = 1, 2, \dots, J$ ,  $t = 1, 2, \dots, n$ , the total number of observations in the sample is given by

$$N = \prod_{j=1}^J \sum_{t=1}^n n_{jt}. \quad (2)$$

---

<sup>2</sup>More precisely, in the empirical application we select option prices from a small window of time, usually 5 to 10 minutes, prior to 3.00pm on each trading day in the estimation sample. Note that although there is some limited variation in the synchronously recorded spot prices during this time period, we continue to use the notation  $S_t$  to denote any spot price recorded during this period on day  $t$ .

The model specified for the  $N$  observed option prices is

$$C_{ijt} = b_{0j} + b_{1j}q(z_{ijt}, \theta) + \sum_{l=1}^4 d_{lj}D_l + \sum_{g=1}^G \rho_{gj}C_{ij(t-g)} + \sigma_j u_{ijt}, \quad (3)$$

$$u_{ijt} \sim N(0, 1) \text{ for all } i = 1, 2, \dots, n_{jt}; j = 1, 2, \dots, J; t = 1, 2, \dots, n. \quad (4)$$

The function  $q(z_{ijt}, \theta)$  in (3) represents the theoretical option price, which is conditional on the assumed distribution of the returns process. As the pricing of the option involves the evaluation of an expectation with respect to the risk-neutral distribution of the underlying asset,  $q(\cdot, \cdot)$  is a function of the parameters which characterize that distribution, denoted by  $\theta$ , in addition to being a function of the vector of observable factors,  $z_{ijt} = (r_t, K_{ij}, \tau_{ij}, S_t)'$ , with  $\tau_{ij}$  representing the maturity of the  $ij$ th option contract and  $r_t$  the risk-free rate of return prevailing on day  $t$ .

The model in (3) allows an observed option price to deviate from the theoretical price in a manner which differs across moneyness group. Specifically, the intercept  $b_{0j}$ , slope  $b_{1j}$  and variance  $\sigma_j^2$  of the model for  $C_{ijt}$  are permitted to vary with  $j$ . In particular, allowance for heteroscedasticity across moneyness groups is necessary as a consequence of the large variation in the magnitude of prices across the moneyness spectrum, a feature that translates into variation across  $j$  in the magnitude of the variance of pricing errors. Dummy variables are also included to capture “day-of-the-week” effects in the option market,  $D_l$ ,  $l = 1, 2, 3, 4$ , where Friday corresponds to  $D_l = 0$  for all  $l$ . The coefficients of the dummy variables,  $d_{lj}$ , are also allowed to vary with  $j$ . The symbol  $C_{ij(t-g)}$  denotes the option price on day  $t - g$  of the  $i$ th contract in moneyness group  $j$ ,  $g = 1, 2, \dots, G$ , for a maximum of  $G$  lags. The lagged dependent variables are included in order to capture correlation across time in pricing errors. With each lagged variable being assigned a group specific coefficient,  $\rho_{gj}$ , the model allows for variation across moneyness groups in the degree of serial correlation in the pricing errors.

The coefficients to be estimated for each moneyness group may be grouped together by moneyness group, and denoted by  $\beta_j = (b_{0j}, b_{1j}, d_{1j}, d_{2j}, d_{3j}, d_{4j}, \rho_{1j}, \dots, \rho_{Gj})'$ , for  $j = 1, \dots, J$ , with  $\beta = (\beta_1', \beta_2', \dots, \beta_J)'$ . The variances associated with each moneyness group may also be grouped as  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_J^2)$ . Further defining  $c_j$  as the  $(N_j \times 1)$  vector of observed options prices for moneyness group  $j$ , ordered by day within the group, with  $N_j = \sum_{t=1}^n n_{jt}$ , the joint density function for  $c = (c_1', c_2', \dots, c_J)'$  is

$$p(c|\Sigma, \beta, \theta) = (2\pi)^{-N/2} \prod_{j=1}^J \sigma_j^{-N_j} \exp \left\{ -\frac{1}{2\sigma_j^2} (c_j - X_j(\theta)\beta_j)' (c_j - X_j(\theta)\beta_j) \right\}, \quad (5)$$

where  $X_j(\theta)$  is an  $(N_j \times L)$  matrix containing the observations on the  $L = 6 + G$  regressors, for moneyness group  $j$ , again ordered by day within the group. The second column of  $X_j(\theta)$  contains the  $N_j$  observations on the theoretical option prices of the contracts in group  $j$ ,  $q(z_{ijt}, \theta)$ . It is via the dependence of  $q(\cdot, \cdot)$  on  $\theta$  that each regressor matrix  $X_j(\theta)$  depends on  $\theta$ . The density in (5) is conditional on initial values for the lagged option prices which appear on the right hand side of (3).<sup>3</sup> Assuming a joint prior for  $\beta$  and  $\Sigma$  of the form

$$p(\beta, \Sigma) \propto \prod_{j=1}^J \sigma_j^{-2}, \quad (6)$$

and imposing a priori independence between  $(\beta, \Sigma)$  and  $\theta$ , the joint posterior for  $\theta$  can be derived as

$$p(\theta|c) \propto \prod_{j=1}^J [X_j(\theta)'X_j(\theta)]^{-1/2} \mathbf{b}_j^{-(N_j-L)} \times p(\theta), \quad (7)$$

where  $p(\theta)$  denotes the prior on  $\theta$ ,  $\mathbf{b}_j^2 = c_j - X_j(\theta)\mathbf{b}_j$ ,  $c_j - X_j(\theta)\mathbf{b}_j / (N_j - L)$  and  $\mathbf{b}_j = [X_j(\theta)'X_j(\theta)]^{-1} X_j(\theta)'c_j$ .

Given the nonstandard nature of (7), which obtains even for the simplest case of the *BS* model, numerical procedures are required in order to produce all posterior quantities of interest. Details of these procedures are provided in Section 4.<sup>4</sup>

### 3 Alternative Option Pricing Models

The evaluation of the option price in (1) and hence the specification of the theoretical option price,  $q(z_{ijt}, \theta)$ , in (3), requires knowledge of the generating process of the spot price  $S_t$ . The assumption underlying the *BS* option pricing model is that returns are normally distributed, with the volatility of returns being constant over the life of the option contract. As is now an established empirical fact, these assumptions do not tally with the observed distributional features of returns, with conditional skewness, leptokurtosis and time-varying volatility being stylized features of most returns data; see Bollerslev, Chou and Kroner (1992) for a review of the relevant literature. As has also been widely documented, *BS*

<sup>3</sup>For notational convenience we do not make explicit the dependence of the joint density for  $c$  on the values of all observable components on the right hand side of equation (3). We also omit these components in the description of all posterior densities.

<sup>4</sup>To rule out arbitrage, the distribution of  $C_{ijt}$  should be truncated from below at  $lb_{ijt} = \max\{0, S_t - e^{-r\tau_{ij}} K_{ij}\}$ ; see Hull (2000). However, the incorporation of this truncation in the likelihood function means that  $(\beta, \Sigma)$  cannot be integrated out analytically. As we wish to minimize the numerical burden associated with the methodology, we choose to omit the truncation at the estimation stage. Note however that in the empirical application we do filter the data according to the lower bound, as well as truncate the predictive densities appropriately in the out-of-sample analysis.



implied volatilities are not constant across strike prices or maturity. Specifically, implied volatility ‘smiles’ or ‘smirks’ across strike prices which, in turn, vary in intensity depending on the time to expiration, have become a stylized fact in empirical work on option prices. Such patterns have been shown to be evidence of implied returns models which deviate from the specifications of the *BS* model; see, for example, Corrado and Su (1997), Hafner and Herwartz (2001) and Lim, Martin and Martin (2002a).

In this section the assumptions which underlie the *BS* model are relaxed, with the distributional frameworks of Lye and Martin (1993, 1994) and Fernandez and Steel (1998) being combined to produce a general model for returns which accommodates both conditional leptokurtosis and skewness. To allow for time-varying volatility over the life of the option, the distributional framework is augmented with a *GARCH*(1,1) model.<sup>5</sup> To price options under this more general specification the risk-neutralization approach of Duan (1995, 1999) is adopted.

### 3.1 Risk-Neutral Specifications

Consider the following empirical model for the continuously compounded return over the small time interval  $\Delta t$ ,

$$\ln S_{t+\Delta t} - \ln S_t = (\mu_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t}e_{t+\Delta t}, \quad (8)$$

where  $\mu_{t+\Delta t}$  is the conditional mean of the return,  $e_{t+\Delta t}$  is a standardized error term and  $\sigma_{t+\Delta t}$  is the annualized conditional volatility of returns. The conditional variance is assumed to follow a *GARCH*(1,1) process,

$$\sigma_{t+\Delta t}^2 = \alpha/\Delta t + \delta\sigma_t^2 e_t^2 + \omega\sigma_t^2, \quad (9)$$

with

$$\alpha > 0; \delta, \omega \geq 0; \delta + \omega < 1.$$

Given the discrete time nature of the model in (8) and (9), the Duan (1995, 1999) approach of using an equilibrium model to specify a local risk-neutral valuation measure, is adopted. In the case where  $e_t$  in (8) is conditionally normal, the (local) risk-neutral process for returns is defined as

$$\ln S_{t+\Delta t} - \ln S_t = (\mu_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t}(z_{t+\Delta t} - \lambda_{t+\Delta t}^N), \quad (10)$$

---

<sup>5</sup>The *GARCH*(1,1) model represents an omnibus model of volatility. More general volatility models which contain asymmetries and longer memory characteristics could be entertained; see, for example, Bauwens and Lubrano (2001) and Bollerslev and Mikkelsen (1999). However, use of these models would increase the number of parameters to be estimated, thereby raising the computational complexity of the Bayesian approach adopted in this paper. Computational issues are discussed in Section 4.

where  $z_{t+\Delta t}$  is the risk-neutral standard normal innovation and  $\lambda_{t+\Delta t}^N$  is a risk premium given by

$$\lambda_{t+\Delta t}^N = \sqrt{\Delta t}(\mu_{t+\Delta t} - r_{t+\Delta t})/\sigma_{t+\Delta t}, \quad (11)$$

where  $r_{t+\Delta t}$  is the risk-free rate of return. The superscript ‘ $N$ ’ in (11) is used to highlight the fact that the risk premium in (11) relates specifically to a normal innovation term. Substitution of (11) in (10) produces the following representation of the risk-neutral process,

$$\ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t}z_{t+\Delta t}. \quad (12)$$

The form of the *GARCH*(1, 1) process under local risk-neutralization is then

$$\sigma_{t+\Delta t}^2 = \alpha/\Delta t + \delta\sigma_t^2 + z_t - \lambda_t^N + \omega\sigma_t^2, \quad (13)$$

which produces an unconditional (annualized) variance equal to

$$\frac{\alpha/\Delta t + \delta\Delta t(\mu_t - r_t)^2}{1 - (\delta + \omega)}. \quad (14)$$

That is, local risk-neutralization implies that given  $\delta > 0$ , options are priced under a distribution with a higher unconditional variance than that associated with the objective process in (9). The extent to which the unconditional variance in (14) exceeds that associated with the objective process depends on the deviation between the actual rate of return on the underlying asset,  $\mu_t$ , and the risk free rate of return,  $r_t$ ; see Duan (1995).

In order to allow for an innovation term in (8) which accommodates skewness and leptokurtosis, the appropriate risk-neutral distribution becomes

$$\ln S_{t+\Delta t} - \ln S_t = (\mu_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t}\Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}), \quad (15)$$

where  $\Psi^{-1}$  denotes the function which transforms the normal variate,  $z_{t+\Delta t}$ , into the relevant nonnormal variate and the risk premium  $\lambda_t$ , is now the solution to

$$E[\Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t})|\mathcal{F}_t] = -\lambda_{t+\Delta t}^N, \quad (16)$$

with  $\mathcal{F}_t$  the set of all information up to time  $t$ ; see Duan (1999) and Hafner and Herwartz (2001). The process for  $\sigma_t^2$  under this so-called generalized local risk-neutral valuation, in turn, becomes

$$\sigma_{t+\Delta t}^2 = \alpha/\Delta t + \delta\sigma_t^2 + [\Psi^{-1}(z_t - \lambda_t)]^2 + \omega\sigma_t^2. \quad (17)$$

To implement the risk-neutral adjustments in (15) to (17) requires several steps, each of which needs to occur at each point in the support of the joint posterior density and at each

point in time in the life of the option over which the process is being simulated. The steps are as follows: 1) repeated numerical simulation of the normal variate,  $z_t$ ; 2) transformation to the relevant nonnormal variate; 3) estimation of the expectation in (16) as a sample mean; and 4) numerical solution to (16) over a grid of values for  $\lambda_t$ . A further transformation from normal to nonnormal random variates, as based on the solution for  $\lambda_t$ , is then required in implementing both (15) and (17), again at each point in the parameter space and at each point in (simulated) time. All of these steps are computationally intensive, especially in the context of conducting implicit Bayesian inference.<sup>6</sup>

To circumvent these computational problems rewrite (15) as

$$\begin{aligned} \ln S_{t+\Delta t} - \ln S_t &= (r_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t} \mathbf{h} \Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}) + \lambda_{t+\Delta t}^N \mathbf{i} \\ &= (r_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t}v_{t+\Delta t}, \end{aligned} \quad (18)$$

where

$$v_{t+\Delta t} = \mathbf{h} \Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}) + \lambda_{t+\Delta t}^N \mathbf{i} \quad (19)$$

is the nonnormal risk-neutral random error term, with conditional mean of zero, given (16). This representation of  $v_{t+\Delta t}$  in (18) and (19) suggests that it can be parameterized directly using a standardized nonnormal density. By definition, the parameters of this distribution, which characterize the higher order moments of the conditional distribution of returns, are the risk-neutralized parameters. These parameters, by construction, differ from the empirical analogues. The risk-neutral process for  $\sigma_t^2$  is, in turn, given by

$$\sigma_{t+\Delta t}^2 = \alpha/\Delta t + \delta\sigma_t^2(v_t - \lambda_t^N)^2 + \omega\sigma_t^2. \quad (20)$$

For consistency, the nonnormal distributional specification adopted for  $v_t$  should nest the normal distribution, in which case  $\lambda_{t+\Delta t} = \lambda_{t+\Delta t}^N$ ,  $\Psi^{-1} = I$ ,  $v_{t+\Delta t} = z_{t+\Delta t}$ , and the processes in (18) and (20) collapse respectively to those in (12) and (13).

In the special case when the volatility is restricted to be constant,  $\sigma_t = \sigma$ , but the assumption of nonnormality is maintained for  $v_t$ , the risk-neutral returns process in (18) reduces to

$$\ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5\sigma^2)\Delta t + \sigma\sqrt{\Delta t}v_{t+\Delta t}. \quad (21)$$

Further, with normality and constant volatility, (18) collapses to

$$\ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5\sigma^2)\Delta t + \sigma\sqrt{\Delta t}z_{t+\Delta t}, \quad (22)$$

---

<sup>6</sup>Note that this same point applies to any estimation method in which the option prices themselves are used as the basis for inference. In Duan (1999), Hafner and Herwartz (2001) and Bauwens and Lubrano (2002), in which *GARCH* option models are estimated using these risk adjustments, the computational burden is much less significant as the parameter estimates are extracted from historical returns data.

which is the discrete version of the risk-neutral distribution which underlies the *BS* option price.

### 3.2 Distributional Specifications

The general specification adopted for the distribution of  $v_t$  in (18) combines elements of the nonnormal distributions formulated in Lye and Martin (1993, 1994) and Fernandez and Steel (1998). Denoting by  $w_t$  a random variable with mean  $\mu_w$  and variance  $\sigma_w^2$ , and defining  $v_t$  via

$$w_t = \sigma_w v_t + \mu_w,$$

the approach of Fernandez and Steel is used to define the density function of  $v_t$  as

$$p_f(v_t) = \frac{2}{\gamma + \frac{1}{\gamma}} \sigma_w^{-1/2} f\left(\frac{w_t}{\gamma}\right) I_{[0,\infty)}(w_t) + f(\gamma w_t) I_{(-\infty,0)}(w_t), \quad (23)$$

where  $f(\cdot)$  is defined as a symmetric density function with a single mode at zero and  $I_A(w)$  denotes the indicator function for the set  $A$ . The mean and variance of  $w_t$  are defined respectively as

$$\mu_w = \frac{\tilde{\mathbf{A}}}{\gamma + 1/\gamma} \int_0^\infty 2x f(x) dx$$

and

$$\sigma_w^2 = \frac{\tilde{\mathbf{A}}}{\gamma + 1/\gamma} \int_0^\infty 2x^2 f(x) dx - \mu_w^2.$$

The parameter  $\gamma$  denotes the degree of skewness in the distribution, with  $\gamma > 1$  corresponding to positive skewness,  $\gamma < 1$  corresponding to negative skewness and  $\gamma = 1$  corresponding to symmetry. The density  $p_f(v_t)$  has a mean of zero, with the sign and magnitude of  $\gamma - 1$  determining the sign and magnitude of the mode. The Pearson skewness coefficient associated with the standardized variate  $v_t$ ,

$$skew = E(v_t), \quad (24)$$

can be computed numerically for any given value of  $\gamma$ .

The density in (23) can be used to produce a standardized skewed normal distribution for  $v_t$  when  $f(\cdot)$  defines the normal density function. Alternatively, defining  $f(\cdot)$  as a density function with excess kurtosis, produces a distribution for  $v_t$  with both leptokurtosis and skewness. By setting  $\gamma = 1$ , symmetric normal and leptokurtic distributions for  $v_t$  are retrieved.

Whilst an obvious choice for the leptokurtic  $f(\cdot)$  density is the Student t density, as pointed out by Duan (1999), such a distribution is problematic when the underlying random

variable is a continuously compounded return. Specifically, the assumption of a Student t distribution for the log-differenced spot price implies that neither the simple return nor the spot price at a given point in time, conditional on the previous spot price, has moments. As the numerical approach adopted in this paper involves simulating returns over successive periods,  $\Delta t$ , then estimating the expectation of a function of the spot price at expiry, it is not feasible to define returns as a Student t variate.<sup>7</sup> Instead, we use a subordinate distribution from the generalized exponential family defined in Lye and Martin (1993, 1994) which has excess kurtosis relative to the normal distribution, but with tail behaviour that ensures the existence of all moments for the spot price process. Defining a random variable  $\eta_t$  with mean and variance  $\mu_\eta$  and  $\sigma_\eta^2$  respectively, this density is defined as

$$f(\eta_t) = k^* \left(1 + \frac{\eta_t^2}{\nu}\right)^{-0.5(\nu+1)/2} \exp(-0.5\eta_t^2), \quad (25)$$

where

$$k^* = \int_{-\infty}^{\infty} \left(1 + \frac{\eta_t^2}{\nu}\right)^{-0.5(\nu+1)/2} \exp(-0.5\eta_t^2) d\eta_t$$

is the normalizing constant. The density in (25) is proportional to a product of Student t and normal kernels. Whilst the first term in the product allows for the excess kurtosis for any finite value of  $\nu$ , the second term ensures that the moments of  $\eta_t$  exist for any value of  $\nu$ . It also ensures that the moments of  $S_t$  taken with respect to the density in (25) also exist for any value of  $\nu$ .

We refer to the density in (25) as the Generalized Student t (*GST*) density. In order to define a *GST* density for the standardized variate  $v_t$ , defined by,

$$\eta_t = \sigma_\eta v_t + \mu_\eta,$$

the variance of  $\eta_t$ ,  $\sigma_\eta^2$ , needs to be computed numerically, along with the integrating constant  $k^*$  in (25). The mean of  $\eta_t$ ,  $\mu_\eta$ , is equal to zero. Whilst there is no closed form expression for the kurtosis in the *GST* distribution, an estimate of the kurtosis coefficient,

$$kurt = E(v_t^4), \quad (26)$$

can be computed numerically for any given value of  $\nu$ .

---

<sup>7</sup>On the other hand, if one were to define the return over the full life of the option as Student t, transform this distribution to the implied distribution of the spot price at maturity, then take the expectation with respect to the latter distribution, the expectation is well-defined, at least for sufficient degrees of freedom; see Lim, Martin and Martin (2002a).

## 4 Implicit Bayesian Inference Using S&P500 Option Prices

### 4.1 Detailed Model Specifications

In this section, S&P500 option price data are used to conduct implicit Bayesian inference on a range of alternative models which are nested in the above distributional framework. Associated with the assumption of constant volatility in (18) are four alternative models for returns, corresponding to the alternative specifications for  $f(\cdot)$  and  $\gamma$  in (23): normal,  $GST$ , skewed normal ( $SN$ ) and skewed  $GST$  ( $SGST$ ), denoted respectively by  $M_1$ ,  $M_2$ ,  $M_3$  and  $M_4$ :

$$\begin{array}{ll}
 M_1 : & f(\cdot) \text{ normal}; \quad \gamma = 1; \quad \sigma_t = \sigma \quad v_t \sim N(0, 1) \\
 M_2 : & f(\cdot) GST; \quad \gamma = 1; \quad \sigma_t = \sigma \quad \mu_\eta + \sigma_\eta v_t \sim GST(\mu_\eta, \sigma_\eta^2, \nu) \\
 M_3 : & f(\cdot) \text{ normal}; \quad \gamma \neq 1; \quad \sigma_t = \sigma \quad \mu_w + \sigma_w v_t \sim SN(\mu_w, \sigma_w^2, \gamma) \\
 M_4 : & f(\cdot) GST; \quad \gamma \neq 1; \quad \sigma_t = \sigma \quad \mu_w + \sigma_w [\mu_\eta + \sigma_\eta v_t] \sim SGST(\mu_w, \sigma_w^2, \gamma, \nu).
 \end{array} \tag{27}$$

As model  $M_1$  corresponds to the discrete time version of the returns model which underlies the  $BS$  option price, we subsequently refer to  $M_1$  as the  $BS$  model. Model  $M_2$  specifies  $v_t$  as  $GST(0, 1, \nu)$ , thereby accommodating excess kurtosis. Model  $M_3$  allows for skewness in returns, whilst model  $M_4$  allows for both leptokurtosis and skewness.

Augmentation of the returns model to cater for the variance structure in (20) leads to additional alternative models, in which the conditional variance is time-varying and the conditional distribution for returns is assumed respectively to be normal,  $GST$ , skewed normal and  $SGST$ . In order to retain parsimony, certain restrictions are placed on the parameterization of the  $GARCH$  models. First, the intercept parameter  $\alpha$  in (20) is set to the value required to equate the risk-neutral unconditional mean of the variance with an average of the estimates of  $\sigma^2$  in the constant volatility models. Secondly, the  $GARCH$ -based models with nonnormal conditional distributions are estimated with the distributional parameters fixed at certain values. Specifically, the models which accommodate excess kurtosis in the distribution of  $v_t$  are estimated with  $\nu$  set to 1.0 and 5.0 respectively. The values of  $\nu$  are chosen so as to produce a continuum of kurtosis behaviour in the conditional distribution of  $v_t$ , ranging from kurtosis of 3 associated with conditional normality, followed by kurtosis of 3.233 associated with  $\nu = 5.0$ , through to kurtosis of 3.624 associated with  $\nu = 1.0$ . In addition, the maximum degree of kurtosis allowed in the conditional distributions of the  $GARCH$  models is deliberately set to be lower than that estimated in the corresponding constant volatility models, as the  $GARCH$  process itself models some of the kurtosis in the unconditional distribution. The model which specifies  $GARCH$  with conditional skewness ( $M_8$ ) is estimated with  $\gamma$  set to 0.85. This value of  $\gamma$  corresponds to a skewness coefficient of

-0.341 and is chosen to reflect the degree of skewness estimated for the corresponding model with constant volatility ( $M_3$ ). The degree of skewness specified for the *GARCH* models with the *SGST* conditional distributions also matches that estimated for the corresponding constant volatility models ( $M_9$  and  $M_{10}$ ) respectively.<sup>8</sup> In total then, six *GARCH* models are estimated, denoted respectively by  $M_5$ ,  $M_6$ ,  $M_7$ ,  $M_8$ ,  $M_9$  and  $M_{10}$  :

$$\begin{aligned}
M_5 : & \quad f(\cdot) \text{ normal}; & \quad \gamma = 1; & \quad \sigma_t & & \quad v_t \sim N(0, 1) \\
M_6 : & \quad f(\cdot) \text{ GST}; \nu = 5 & \quad \gamma = 1; & \quad \sigma_t & & \quad \mu_\eta + \sigma_\eta v_t \sim \text{GST}(\mu_\eta, \sigma_\eta^2, \nu) \\
M_7 : & \quad f(\cdot) \text{ GST}; \nu = 1 & \quad \gamma = 1; & \quad \sigma_t & & \quad \mu_\eta + \sigma_\eta v_t \sim \text{GST}(\mu_\eta, \sigma_\eta^2, \nu) \\
M_8 : & \quad f(\cdot) \text{ normal}; & \quad \gamma = 0.85; & \quad \sigma_t & & \quad \mu_w + \sigma_w v_t \sim \text{SN}(\mu_w, \sigma_w^2, \gamma) \\
M_9 : & \quad f(\cdot) \text{ GST}; \nu = 5 & \quad \gamma = 0.80; & \quad \sigma_t & \quad \mu_w + \sigma_w [\mu_\eta + \sigma_\eta v_t] & \sim \text{SGST}(\mu_w, \sigma_w^2, \gamma, \nu). \\
M_{10} : & \quad f(\cdot) \text{ GST}; \nu = 1 & \quad \gamma = 0.80; & \quad \sigma_t & \quad \mu_w + \sigma_w [\mu_\eta + \sigma_\eta v_t] & \sim \text{SGST}(\mu_w, \sigma_w^2, \gamma, \nu).
\end{aligned} \tag{28}$$

Models  $M_1$  to  $M_{10}$  all imply a different functional form for the theoretical option price,  $q(z_{ijt}, \theta)$ , in (3), as well as a different specification for the parameter vector,  $\theta$ . As noted earlier, for all models other than  $M_1$ ,  $q(z_{ijt}, \theta)$  does not have a closed-form solution. For the models  $M_2$  to  $M_4$  the approach adopted is to simulate (21) over the life of the contract, with the innovations drawn from the relevant nonnormal distribution in (27). For each of these models, simulation of the relevant process for returns is repeated  $h$  times, producing  $S_T^{(l)}$ ,  $l = 1, 2, \dots, h$ , and the expectation in (1) approximated by the sample mean of  $e^{-r_t \tau_{ij}} \max(S_T^{(l)} - K_{ij}, 0)$ . Both antithetic and control variates are used to reduce the simulation error, with the analytical *BS* option price used as the control variate. For the six time-varying volatility models,  $M_5$  to  $M_{10}$ , the processes in (18) and (20) are simulated over the life of the option. For a general discussion of this simulation-based approach to the pricing of options see [Gourieroux and Monfort \(1994\)](#) and for some recent applications, see [Bollerslev and Mikkelsen \(1999\)](#), [Duan \(1999\)](#), [Hafner and Herwartz \(2001\)](#) and [Bauwens and Lubrano \(2002\)](#).

In the simulation of all relevant processes,  $\Delta t = 1/365$ , thereby representing one day. As such, all estimated parameters can be interpreted as the option-implied estimates associated with daily returns. The exception to this is the volatility parameter in the constant volatility models which, following convention, is reported as an annualized figure.

---

<sup>8</sup>Since the *GARCH* model does not accommodate asymmetry in returns, it is legitimate to specify a degree of skewness in the associated conditional distribution which is equivalent to that in the unconditional distribution of the corresponding constant volatility model.

## 4.2 Data Description

The data are based on bid-ask quotes on call options written on the S&P500 stock index, obtained from the Berkeley Options Database. The quotes relate to options traded during the first 239 trading days of 1995, 3/1/1995 to 15/12/1995, during a period of approximately ten minutes immediately prior to 3.00pm on each day. As noted earlier, this form of data selection was aimed at minimizing the amount of intraday variation in the spot prices recorded synchronously with the option prices. A cross section of approximately 60 prices is selected on each day, with the prices deliberately chosen so as to span the full moneyness spectrum. Defining  $S_t - K_{ij}$  as the intrinsic value of the  $i$ th call option in moneyness group  $j$  priced at time  $t$ , options for which  $S_t/K_{ij} \in (0.98, 1.04)$  are categorized as at-the-money (ATM), those for which  $S_t/K_{ij} \leq 0.98$ , as out-of-the-money (OTM), and those for which  $S_t/K_{ij} \geq 1.04$ , as in-the-money (ITM); see Bakshi, Cao and Chen (1997). The options in the sample can be classified as short-term as maturity lengths range from approximately one week to approximately one and a half months. Each record in the dataset comprises the bid-ask quote, the synchronously recorded spot price of the index, the time at which the quote was recorded, and the strike price. As dividends are paid on the S&P500 index, in the option price formulae the current spot price,  $S_t$ , is replaced by the dividend-exclusive spot price,  $S_t e^{-D\tau_{ij}}$ , where  $D = 0.026$  is the average annualized dividend rate paid over the life of the option, with  $D$  estimated from dividend data for 1995 and 1996 obtained from Standard and Poors. The risk-free rate  $r_t$  is set at the average annualized three month bond rate for 1995,  $r = 0.057$ . A constant value of  $r$ , rather than a time series of daily values, is adopted for computational convenience and is justified by the minimal amount of variation in the three month bond rate over 1995. Filtering the data according to the no-arbitrage lower bound of  $lb = \max\{0, S_t e^{-D\tau_{ij}} - e^{r_t\tau_{ij}} K_{ij}\}$  leaves 8968 observations in the estimation sample, for which the main characteristics are summarized in Panel A in Table 1.

The out-of-sample performance of the alternative models is based on option price quote data recorded in the few minutes before 3.00pm on each day from 18/12/1995 to 22/12/1995, with the same dividend adjustment and lower bound filtering as is applied to the estimation dataset, having been applied to the hold-out sample. A total of 984 option prices are used to assess the out-of-sample performance of the models. The characteristics of this dataset are summarized in Table 1, Panel B. The most important difference between the estimation and hold-out sample is the lack of any OTM options in the latter. In addition, even in the ATM range, the out-of-sample options tend toward the higher end of that range, with the average price and bid-ask spread being larger as a consequence, than the corresponding



Table 1:  
S&P500 Option Price Dataset

Moneyness ( $S_t/K_{ij}$ )	Average Market Price	Average Bid-Ask Spread	No. of Prices
Panel A: Estimation Dataset: 3/1/1995 to 15/12/1995			
OTM: $< 0.98$	\$0.72	\$0.12	440
ATM: $0.98 - 1.04$	\$10.90	\$0.50	2209
ITM: $\geq 1.04$	\$68.99	\$0.97	6319
		Total	8968
Panel B: Out of Sample Dataset: 18/12/1995 to 22/12/1995			
OTM: $< 0.98$	n.a. <sup>(a)</sup>	n.a. <sup>(a)</sup>	0
ATM: $0.98 - 1.04$	\$20.61	\$0.87	166
ITM: $\geq 1.03$	\$70.38	\$1.00	818
		Total	984

(a) Not applicable.

figures in the estimation sample. The average prices and bid-ask spreads for both sets of ITM options are very similar.

### 4.3 Priors

The Bayesian analysis is based on a noninformative prior for the constant volatility parameter,  $\sigma$ , and informative priors for the degrees of freedom and skewness parameters,  $\nu$  and  $\gamma$  respectively. A-priori independence between all parameters is imposed. The standard noninformative prior is used for  $\sigma$ ,  $p(\sigma) \propto 1/\sigma$ , despite the fact that its rationale as a Jeffreys prior no longer holds, given the form of the likelihood function in (5). By specifying the same prior for  $\sigma$  in all of  $M_1$  to  $M_4$ , the Bayes factors used for all pairs of these models are unaffected by the fact that this prior is improper. An inverted gamma prior is specified for  $\nu$ , with  $E(\nu) = 1.76$  and  $var(\nu) = 197.89$ . The prior is calibrated so as to match approximately the location of the posterior density for  $\nu$  based on Bayesian estimation of a *GST* model for 1995 daily returns data, but with the variance of the prior being several-fold larger than the variance of the returns posterior. A normal prior is specified for  $\gamma$ , with  $E(\gamma) = 1.0$  and  $var(\gamma) = 1.0$ . Again, the prior is calibrated to match the location of the posterior density for  $\gamma$  estimated from the 1995 daily returns data, but with the variance of the prior specified to be much larger.<sup>9</sup> For the *GARCH* models, a uniform prior is placed on the joint space of  $\delta$  and  $\omega$ , bounded by  $\delta \geq 0$ ,  $\omega \geq 0$  and  $\delta + \omega < 1$ .

### 4.4 Implicit Posterior Density Estimates

The first step in the implicit analysis is to produce estimates of the marginal posterior distributions for the parameters of the alternative models. Defining  $\theta_k$  as the parameter vector associated with model  $M_k$ ,  $k = 1, 2 \dots 10$ , the joint posterior for  $\theta_k$ ,  $p(\theta_k|c)$ , is given by (7), with  $c$  denoting the vector of 9864 option prices observed during the estimation sample period. For all ten models,  $p(\theta_k|c)$  is normalized and marginal posteriors produced via deterministic numerical integration. Independent samples from each  $p(\theta_k|c)$  are produced using the inverse cumulative distribution function technique. This approach is feasible due to the highly parsimonious nature of the distributional models, in conjunction with the restrictions placed on the parameters of the *GARCH* models,  $M_5$  to  $M_{10}$ .<sup>10</sup> The advantage

<sup>9</sup>Note that the Bayes factors related to the models in which  $\nu$  and  $\gamma$  feature are well-defined only when proper priors are specified for these parameters. One way of avoiding the usual arbitrariness associated with the prior specification is to use the returns data to determine their essential form; see also Jacquier and Jarrow (2000).

<sup>10</sup>In evaluating  $\lambda_i^N$  in (11), a constant mean is assumed for the empirical returns distribution, whereby  $\mu_t$  is replaced by the sample mean of returns for 1995.

of this numerical approach is that the results produced are essentially exact, with none of the convergence issues which would be associated with a Markov Chain sampling algorithm. This is particularly important in the present context in which the theoretical option prices themselves, for all models other than  $M_1$ , need to be computed using computationally intensive numerical simulation. That is, it would not be computationally feasible to produce the number of Markov Chain iterates required to establish convergence, in combination with the Monte Carlo-based estimation of the theoretical option prices.

In Table 2, the mean, mode and approximate 95% Highest Posterior Density (HPD) intervals are reported for each parameter in the ten models estimated<sup>11</sup>. The first thing to note is the similarity across the four constant volatility models,  $M_1$  to  $M_4$ , of the point estimates of volatility. The modal estimate of  $\sigma$  varies only between 0.115 for  $M_1$ ,  $M_3$  and  $M_4$  and 0.125 for  $M_2$ . As the densities are essentially symmetric, the mean estimates are equivalent to the modal estimates, with the degree of dispersion in the densities also equal across models.

The modal point estimates of the degrees of freedom parameter,  $\nu$ , in both  $M_2$  and  $M_4$ , are equal to 0.85, with the mean values only slightly higher, at 0.934 and 0.919 respectively. These three point estimates of  $\nu$  imply (estimates of) the kurtosis coefficient in (26) of 3.674, 3.645 and 3.650 respectively. Remembering that, by construction, both  $\nu$  and  $\gamma$  are interpreted as distributional parameters for implicit daily returns distributions, these kurtosis values are representative of returns distributions with a moderate degree of excess kurtosis. The 95% interval estimates cover values for  $\nu$  which translate into kurtosis values which all exceed the value of 3 associated with normality. The modal estimates of the skewness parameter,  $\gamma$ , in  $M_3$  and  $M_4$ , are 0.85 and 0.80 respectively, thereby indicating negative skewness in the implicit daily returns distribution, with (estimates of) the skewness coefficient in (24) of  $-0.253$  and  $-0.341$  respectively. For  $M_3$  in particular, however, the distribution of  $\gamma$  is positively skewed, with a mean estimate close to unity. Moreover, the 95% intervals for  $\gamma$  in both models are very wide, easily covering values for  $\gamma$  which imply either symmetry ( $\gamma = 1$ ) or positive skewness ( $\gamma > 1$ ), in addition to values implying negative skewness ( $\gamma < 1$ ). Some of these numerical results are illustrated graphically in Figure 1, in which the marginal densities for the distributional parameters in models  $M_2$  and  $M_3$  are reproduced. For  $M_2$  the posterior density of the estimated kurtosis coefficient is also presented, providing clear evidence of option-implied excess kurtosis. For  $M_3$  the

---

<sup>11</sup>An HPD interval is an interval with the specified probability coverage, whose inner density ordinates are not exceeded by any density ordinates outside the interval. The reported intervals have a coverage which is as close to the nominal coverage as possible given the discrete grid defined for each parameter.

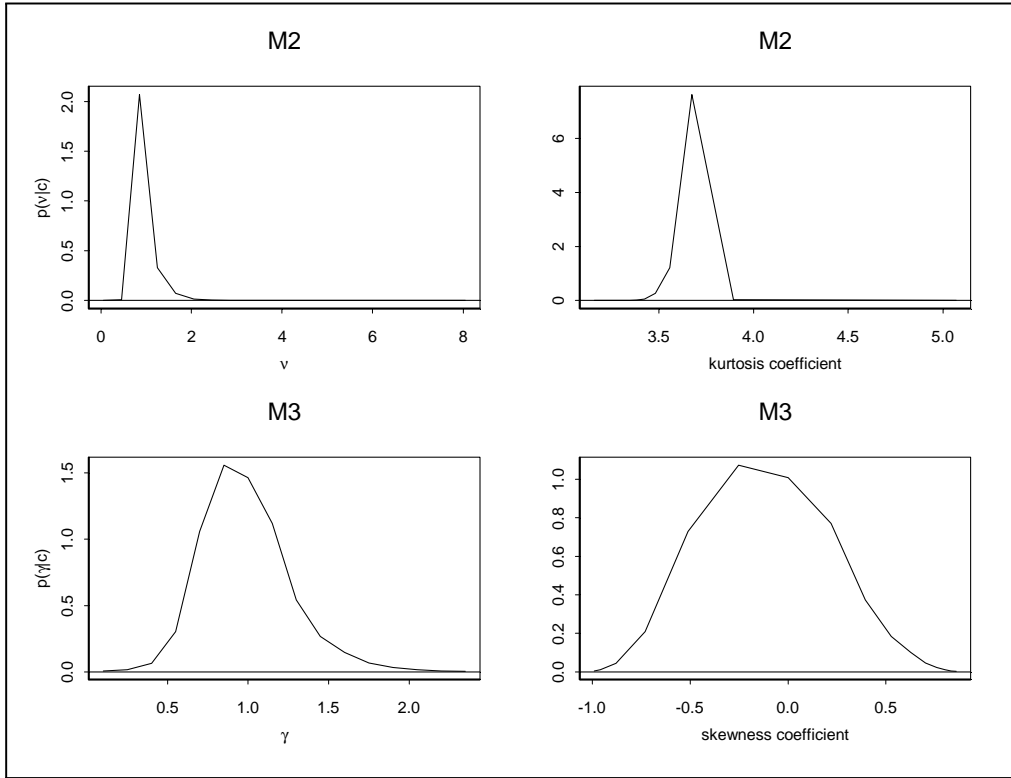


Figure 1: Implicit Marginal Posteriors for Selected Parameters.

posterior density of the estimated Pearson skewness coefficient is presented in addition to the posterior density for the skewness parameter  $\gamma$ , making clear the fact that all forms of skewness are given high posterior weight, despite the negative mode for  $\gamma$ .

For all six time-varying volatility models,  $M_5$  to  $M_{10}$ , the option-implied persistence in daily volatility,  $\theta + \delta$ , is low in comparison with typical returns-based estimates, ranging from 0.8 to 0.84 in terms of point estimates. In addition, the small values estimated for  $\delta$  indicate that the volatility process evolves relatively smoothly over the life of the option.<sup>12</sup> By construction, the long-run volatility is held fixed at an annualized value of 0.12 in all cases.

## 4.5 Model Rankings

### 4.5.1 Implicit Model Probabilities

Implicit model probabilities are derived from the posterior odds ratios, constructed for each model,  $M_2, M_3, \dots, M_{10}$ , relative to a reference model,  $M_1$ . Defining  $P(M_k|c)$  as the

<sup>12</sup>Using the EVIEWS program to estimate a  $GARCH(1,1)$  models for daily returns on the S&P500 index for the period 1994 to 1997, estimates similar to those reported in Table 2 are obtained.

Table 2:  
Implicit Marginal Posterior Densities<sup>(a)</sup>

Model		Parameter	Mode	Mean	95% HPD Interval
$M_1$		$\sigma$	0.115	0.115	(0.106, 0.124)
$M_2$		$\sigma$	0.125	0.125	(0.116, 0.134)
		$\nu$	0.850	0.934	(0.450, 1.650)
$M_3$		$\sigma$	0.115	0.115	(0.106, 0.124)
		$\gamma$	0.850	0.986	(0.400, 1.600)
$M_4$		$\sigma$	0.115	0.115	(0.106, 0.124)
		$\nu$	0.850	0.919	(0.250, 2.100)
		$\gamma$	0.800	0.891	(0.650, 1.150)
$M_5$		$\delta$	0.030	0.031	(0.022, 0.038)
		$\omega$	0.810	0.810	(0.802, 0.818)
$M_6$	$v = 5.0$	$\delta$	0.030	0.031	(0.022, 0.038)
		$\omega$	0.810	0.810	(0.802, 0.818)
$M_7$	$v = 1.0$	$\delta$	0.031	0.031	(0.022, 0.038)
		$\omega$	0.810	0.810	(0.802, 0.818)
$M_8$	$\gamma = 0.85$	$\delta$	0.040	0.040	(0.031, 0.049)
		$\omega$	0.760	0.076	(0.751, 0.769)
$M_9$	$v = 5.0; \gamma = 0.80$	$\delta$	0.030	0.030	(0.022, 0.038)
		$\omega$	0.780	0.780	(0.771, 0.789)
$M_{10}$	$v = 1.0; \gamma = 0.80$	$\delta$	0.030	0.030	(0.022, 0.038)
		$\omega$	0.780	0.780	(0.772, 0.788)

(a) By convention  $\sigma$  is reported as an annualized quantity. The distributional parameters  $\nu$  and  $\gamma$  relate to daily returns, whilst the sum of the *GARCH* parameters,  $\delta$  and  $\omega$ , measures daily persistence in volatility.

posterior probability of  $M_k$ , the posterior odds ratio for  $M_k$  versus  $M_1$  is given by

$$\begin{aligned} \frac{P(M_k|c)}{P(M_1|c)} &= \frac{P(M_k)}{P(M_1)} \times \frac{p(c|M_k)}{p(c|M_1)} \\ &= \text{Prior Odds} \times \text{Bayes Factor}, \end{aligned} \quad (29)$$

for  $k = 2, 3, \dots, 10$ , where

$$p(c|M_k) = \int_{\Sigma} \int_{\beta} \int_{\theta_k} L(\Sigma, \beta, \theta_k|M_k) p(\Sigma, \beta, \theta_k|M_k) d\Sigma d\beta d\theta_k, \quad (30)$$

is the marginal likelihood of  $M_k$ , with  $L(\Sigma, \beta, \theta_k|M_k)$  and  $p(\Sigma, \beta, \theta_k|M_k)$  respectively denoting the likelihood and prior under  $M_k$ . The model probabilities are calculated by solving the nine ratios in (29) subject to the normalization

$$\sum_{k=1}^{10} P(M_k|c) = 1. \quad (31)$$

The models are then ranked as a posteriori most probable to least probable according to the size of the probabilities. As  $\Sigma$  and  $\beta$  can be integrated out analytically, the marginal likelihood for model  $M_k$  reduces to

$$p(c|M_k) = h \int_{\theta_k} L(\theta_k|M_k) p(\theta_k|M_k) d\theta_k, \quad (32)$$

where  $h$  is a constant which is independent of the specification of  $M_k$ . The integral in (32) is that which is computed in the numerical normalization of the posterior density for  $\theta_k$  in (7). Hence, the marginal likelihood for each model arises as a natural by-product of the numerical approach adopted, rather than requiring additional computation. Computation of the Bayes factors and implicit probabilities then follows.

Table 3 provides the estimated Bayes factors for the ten models  $M_1$  to  $M_{10}$ , with  $M_1$  used as the reference model. The final row gives the associated model probabilities, based on equal prior probabilities in (29) for all ten models. There are three notable aspects of the results in Table 3. First, the *GST* model with constant volatility ( $M_2$ ) is assigned all posterior probability (to two decimal places) in the set of ten alternative models. This is completely consistent with the fact that the option prices have factored in distributional estimates which imply excess kurtosis, as indicated by the results reported in Table 2. Secondly, despite the dominance of the *GST* model, there is a clear hierarchy amongst the other three constant volatility models, namely  $M_1$  is favoured over  $M_4$ , which is, in turn, favoured over  $M_3$ . That is, amongst the four constant volatility models, the *BS* model is ranked second according to posterior probability weight. Thirdly, all six *GARCH*-based

Table 3:  
 Implicit Bayes Factors and Model Probabilities.  
 Entry  $(i, j)$  Indicates the Bayes Factor  
 in Favour of  $M_j$  Versus  $M_i$

	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
$M_1$	1.00	31400	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$M_2$		1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$M_3$			1.00	1200	0.00	0.00	0.00	0.00	0.00	0.00
$M_4$				1.00	0.00	0.00	0.00	0.00	0.00	0.00
$M_5$					1.00	2050	$8.3E07$	31.30	$8.8E09$	$9.2E25$
$M_6$						1.00	40260	0.00	$4.3E06$	$4.5E22$
$M_7$							1.00	0.00	106	$1.1E18$
$M_8$								1.00	$2.8E08$	$3.0E24$
$M_9$									1.00	$1.1E16$
$M_{10}$										1.00
$P(M_k c)$	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

models are assigned essentially zero probability when ranked against any of the constant volatility models. The dominance of the constant volatility models reflects the low values in the support of the marginal density for  $\delta$  in the *GARCH* specification in (20), which are, in turn, associated with a smoothly evolving volatility process over the life of the option. This results in models  $M_5$  to  $M_{10}$  being effectively overparameterized and, hence, penalized in comparison with the constant volatility models. However, when considered as a separate set, there is a clear ranking across the time-varying volatility models, with the models which impose both excess kurtosis and some negative skewness in the conditional distribution ( $M_9$  and  $M_{10}$ ) favoured most highly, followed by the models with conditional kurtosis only ( $M_6$  and  $M_7$ ), followed in turn by the conditional skewness model ( $M_8$ ), then by the conditional normal model ( $M_5$ ).

#### 4.5.2 Out-of-Sample Fit Performance

For model  $M_k$  with parameter vector  $\theta_k$ , the residual associated with fitting the  $i$ th option price  $C_{ijf}$ , for moneyness group  $j$ , observed on some day  $f$  during the hold-out sample is

defined as

$$\begin{aligned}
res_{ijf} &= C_{ijf} - b_{0j} - b_{1j}q(z_{ijf}, \theta_k) - \sum_{l=1}^{\mathcal{A}} d_{lj}D_l - \sum_{g=1}^{\mathcal{G}} \rho_{gj}C_{ij(f-g)} \\
&= C_{ijf} - x_{ijf}(\theta_k)' \beta_j,
\end{aligned} \tag{33}$$

where  $z_{ijf}$  denotes the option contract specifications associated with  $C_{ijf}$ ,  $x_{ijf}(\theta_k)'$  is a  $(1 \times L)$  vector of observations at time period  $f$  on the  $L = 6 + G$  regressors associated with  $C_{ijf}$ , and  $\beta_j$  is the  $(L \times 1)$  regression vector associated with moneyness group  $j$ . Standard Bayesian distribution theory for a normal linear model yields a multivariate Student  $t$  posterior distribution for  $\beta_j$ , conditional on  $\theta_k$ , with

$$E(\beta_j | \theta_k, c) = \mathfrak{b}_j$$

and

$$var(\beta_j | \theta_k, c) = \mathfrak{b}_j^2 \frac{h}{i-1} X_j(\theta_k)' X_j(\theta_k)^{-1},$$

where  $\mathfrak{b}_j$  and  $\mathfrak{b}_j^2$  are as defined previously in the text. Hence, the posterior distribution for  $res_{ijf}$ , conditional on  $\theta_k$ , is univariate Student  $t$ , with

$$E(res_{ijf} | \theta_k, c) = C_{ijf} - x_{ijf}(\theta_k)' \mathfrak{b}_j \tag{34}$$

and

$$var(res_{ijf} | \theta_k, c) = \mathfrak{b}_j^2 x_{ijf}(\theta_k)' \frac{h}{i-1} X_j(\theta_k)' X_j(\theta_k)^{-1} x_{ijf}(\theta_k). \tag{35}$$

The marginal posterior for  $res_{ijf}$  is thus defined as

$$p(res_{ijf} | c) = \int_{\theta_k} p(res_{ijf} | \theta_k, c) p(\theta_k | c) d\theta_k. \tag{36}$$

As  $p(\theta_k | c)$  is specified numerically over the grid of values for  $\theta_k$  used in the numerical normalization of  $p(\theta_k | c)$ , the integral in (36) can be estimated by taking a weighted sum of Student  $t$  densities, with the weights determined by  $p(\theta_k | c)$ . Given an estimate of  $p(res_{ijf} | c)$ , a 95% HPD interval for  $res_{ijf}$  can be calculated. For any given model  $M_k$  there is a residual interval for each option price in the hold-out sample of 984 prices. The proportion of intervals which cover zero is a measure of how well the model fits out-of-sample, with the best fitting model defined as the model for which this proportion is the highest.

Results are reported in Table 4 both for the two moneyness groups which are represented out-of-sample: ATM and ITM, and for the full out-of-sample dataset. The number of options in these three groups are respectively 166, 818 and 984. Also included in the lower portion of the table, for all three categories of option, are the average sizes of the bid-ask





### 4.5.3 Out-of-Sample Predictive Performance

For model  $M_k$ , the predictive density for option price  $C_{ijf}$  is given by

$$p(C_{ijf}|c) = \int_{\beta_j} \int_{\sigma_j} \int_{\theta_k} p(C_{ijf}|\beta_j, \sigma_j, \theta_k, c) p(\beta_j, \sigma_j|\theta_k, c) p(\theta_k|c) d\beta_j d\sigma_j d\theta_k, \quad (37)$$

where  $p(C_{ijf}|\beta_j, \sigma_j, c, \theta_k)$  is a normal density, given the assumption of a normal distribution for  $u_{ijf}$  in (3). Again, standard Bayesian results enable analytical integration with respect to  $\beta_j$  and  $\sigma_j$  such that

$$p(C_{ijf}|c) = \int_{\theta_k} p(C_{ijf}|\theta_k, c) p(\theta_k|c) d\theta_k, \quad (38)$$

where  $p(C_{ijf}|\theta_k, c)$  is a univariate Student  $t$  density with

$$E(C_{ijf}|\theta_k, c) = x_{ijf}(\theta_k)' \mathbf{b}_j \quad (39)$$

and

$$\text{var}(C_{ijf}|\theta_k, c) = \mathbf{b}_j^2 [1 + x_{ijf}(\theta_k)' (X_j(\theta_k)' X_j(\theta_k))^{-1} x_{ijf}(\theta_k)]. \quad (40)$$

The predictive density in (38) can be estimated as a weighted sum of Student  $t$  densities, with weights given by  $p(\theta_k|c)$ . Truncation of  $p(C_{ijf}|\theta_k, c)$  at the no-arbitrage lower bound is imposed before averaging over the space of  $\theta_k$ . A comparison of (40) with (35) reveals that the Student  $t$  densities used in the mixture which defines the predictive in (38) have a variance which is larger by a factor of  $\mathbf{b}_j^2$  than the variance of the densities used in the construction of the residual function. This result reflects the standard linear regression structure of the model for the option pricing errors in (3) and mimics the classical prediction results associated with that model.

The estimated predictive density is used to rank the predictive performance of the models in several different ways. First, it is used to assign a probability to the observed bid-ask spread associated with the option contract for which  $C_{ijf}$  is the market price.<sup>13</sup> This calculation is repeated for all option contracts, the predictive probability recorded for model  $M_k$  being the average of all computed probabilities. Second, with the predictive mode taken as a point predictor of  $C_{ijf}$ , the accuracy of each model is assessed in terms of the proportion of predictive modes which fall within the observed bid-ask spreads.<sup>14</sup> The same

<sup>13</sup>With regard to the S&P500 option price data, there is usually only one bid-ask spread associated with a particular option contract, where the specification of that contract includes the current spot price of the index. For some contracts, however, there are several bid-ask spreads quoted. These spreads are averaged over before being used in the probability calculation described in the text.

<sup>14</sup>Note that there is a large literature on the market related factors which influence the bid-ask spreads associated with option prices. In particular, attempts have been made to explain the way in which the spreads vary across different type of option contracts; see, for example George and Longstaff (1993). On the assumption that these factors do not relate to the nature of the underlying returns process, the observed spreads can be treated as given intervals to which the different models assign varying predictive probabilities. This assumption may be questionable however; see, for instance, Cho and Engle (1999).

calculation is performed for the predictive means. Third, the proportion of market prices which fall within the 95% probability interval associated with the estimated predictive, is calculated for each model. As with the fit results, all calculations are performed for ATM and ITM contracts as well as for all 984 contracts in the hold-out sample, with information on the average bid-ask-spreads and the average width of the model-specific intervals also included. The results for the three different contract groupings are reported in Tables 5, 6 and 7 respectively.

As is the case with the fit results, the predictive results indicate that the constant volatility models with nonnormal distributional specifications,  $M_2$ ,  $M_3$  and  $M_4$ , have the best performance out-of-sample. This is the case for both the ATM and ITM options. In terms of the proportion of times that the point predictors, the predictive mean and mode, fall in the bid-ask spread, the *BS* model is the next best performer, whilst the *GARCH* models tend to have a slightly better predictive performance than the *BS* model in terms of the observed price falling within the 95% predictive interval. It should be noted, however, that the average width of this interval, in the case of the *GARCH* models, tends to be larger than the average width associated with the *BS* intervals, at least for the ITM options. The *BS* and *GARCH* models ascribe very similar probabilities to the observed bid-ask spreads, all of which are lower than the corresponding probabilities ascribed by the non-*BS* constant volatility models. Focussing on the overall results for all out-of-sample options, as reported in Table 7, the average probability ascribed to the bid-ask spread ranges from 31.7% for  $M_8$  and  $M_9$  to 33.9% for  $M_2$ . If the predictive mode is used as a point predictor of the option price, the results in Table 7 show that the probability of predicting an option price within the observed spread ranges from 20.5% for  $M_8$  to 26.9% for  $M_4$ . The predictive mean serves as a more accurate point predictor, with the probability of it falling within the observed spread ranging from 26.6% for  $M_9$  to 32.6% for  $M_4$ . The 95% predictive interval covers the observed market price approximately 70% of the time for all models, with  $M_4$  again having the best performance overall according to this criterion. Note however, that whilst the coverage of the predictive intervals appears to be reasonable for all models, the average width of the intervals does exceed the average width of the bid-ask spread, and, hence, could be viewed as being too broad an interval to be useful from a practical point of view.

Table 5:  
 Predictive Performance of the Different Models (ATM Options)

Predictive Criterion	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
Prob( $ba$ ) <sup>(a)</sup>	0.264	0.323	0.326	0.321	0.277	0.278	0.280	0.278	0.276	0.276
Mode in $ba$ <sup>(b)</sup>	0.300	0.299	0.312	0.314	0.290	0.290	0.288	0.284	0.278	0.283
Mean in $ba$ <sup>(b)</sup>	0.312	0.333	0.340	0.346	0.308	0.312	0.312	0.312	0.308	0.312
Price in 95% I <sup>(b),(c)</sup>	0.518	0.591	0.613	0.613	0.549	0.549	0.553	0.541	0.541	0.541
Average Bid-Ask Spread	Average Width of 95% Prediction Intervals									
	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
\$0.87	\$1.85	\$1.81	\$1.81	\$1.80	\$1.96	\$1.96	\$1.96	\$1.94	\$1.93	\$1.93

- (a)  $ba$  = the bid-ask spread. The figures reported in this line are the average of the 166 predictive probabilities calculated for each model.
- (b) All figures reported are proportions of 166.
- (c) The 95% Interval is the interval which excludes 2.5% in the lower and upper tails of the predictive distribution. This interval equals the 95% HPD interval only for those predictives which are symmetric around a single mode.

Table 6:  
 Predictive Performance of the Different Models (ITM Options)

Predictive Criterion	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
Prob( $ba$ ) <sup>(a)</sup>	0.325	0.343	0.340	0.340	0.327	0.328	0.329	0.322	0.323	0.326
Mode in $ba$ <sup>(b)</sup>	0.225	0.249	0.258	0.260	0.198	0.204	0.211	0.183	0.188	0.198
Mean in $ba$ <sup>(b)</sup>	0.281	0.321	0.320	0.322	0.265	0.264	0.267	0.255	0.252	0.259
Price in 95% I <sup>(b),(c)</sup>	0.667	0.696	0.697	0.698	0.687	0.692	0.693	0.676	0.677	0.683
Average Bid-Ask Spread	Average Width of 95% Prediction Intervals									
	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
\$1.00	\$1.71	\$1.73	\$1.75	\$1.75	\$1.72	\$1.73	\$1.73	\$1.71	\$1.71	\$1.71

- (a)  $ba$  = the bid-ask spread. The figures reported in this line are the average of the 818 predictive probabilities calculated for each model.
- (b) All figures reported are proportions of 818.
- (c) The 95% Interval is the interval which excludes 2.5% in the lower and upper tails of the predictive distribution. This interval equals the 95% HPD interval only for those predictives which are symmetric around a single mode.

Table 7:  
 Predictive Performance of the Different Models (All Options)

Predictive Criterion	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
Prob( $ba$ ) <sup>(a)</sup>	0.317	0.339	0.337	0.337	0.321	0.322	0.323	0.317	0.317	0.320
Mode in $ba$ <sup>(b)</sup>	0.241	0.256	0.268	0.269	0.218	0.224	0.230	0.205	0.208	0.217
Mean in $ba$ <sup>(b)</sup>	0.290	0.321	0.324	0.326	0.276	0.276	0.278	0.269	0.266	0.272
Price in 95% I <sup>(b),(c)</sup>	0.646	0.679	0.684	0.684	0.668	0.673	0.675	0.658	0.659	0.664
Average Bid-Ask Spread	Average Width of 95% Prediction Intervals									
	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$	$M_7$	$M_8$	$M_9$	$M_{10}$
\$0.98	\$1.73	\$1.74	\$1.76	\$1.76	\$1.77	\$1.77	\$1.77	\$1.75	\$1.75	\$1.75

- (a)  $ba$  = the bid-ask spread. The figures reported in this line are the average of the 984 predictive probabilities calculated for each model.
- (b) All figures reported are proportions of 984.
- (c) The 95% Interval is the interval which excludes 2.5% in the lower and upper tails of the predictive distribution. This interval equals the 95% HPD interval only for those predictives which are symmetric around a single mode.

#### 4.5.4 Hedging Performance

Another measure of the performance of alternative option price models is the extent to which the associated hedging errors deviate from zero. In this paper attention is restricted to delta hedges. The delta for the  $i$ th option price, in moneyness group  $j$ , observed at time  $t$ , based on the assumption that  $M_k$  describes the returns process, is defined as

$$\delta_k = \frac{\partial q(z_{ijt}, \theta_k)}{\partial S_t}. \quad (41)$$

In computing the hedging errors, the portfolio consists of going short in the option and long in the underlying asset by an amount of  $\delta_k$  shares in the asset, and investing the residual,  $C_{ijt} - \delta_k S_t$ , at the risk free interest rate  $r$ . At time  $t + \Delta t$ , the hedging error over a time interval  $\Delta t$ , is given by; see Bakshi, Cao and Chen (1997)

$$H_k = \delta_k (S_{t+\Delta t} - S_t e^{r\Delta t}) - (C_{ij(t+\Delta t)} - C_{ijt} e^{r\Delta t}). \quad (42)$$

The posterior distribution of the hedging error in (42) is derived from the posterior distribution for the parameters of model  $M_k$ , via  $\delta_k$ . In fact, the distribution of  $H_k$  is a simple translation of the distribution of  $\delta_k$ , obtained by recentering this distribution by  $C_{ij(t+\Delta t)} - C_{ijt} e^{r\Delta t}$ , and rescaling it by  $S_{t+\Delta t} - S_t e^{r\Delta t}$ . Thus, the hedging error density,  $p(H_k|c)$ , can be generated by evaluating  $H_k$ , via  $\delta_k$ , at values of  $\theta_k$  in the support of  $p(\theta_k|c)$ , and defining  $p(H_k|c)$  according to the probability weights given by the numerically normalized  $p(\theta_k|c)$ . The model with the hedging error density most closely concentrated around zero is, according to this criterion, the best model.

Two hedge distributions are constructed, based respectively on one-day and five-days ahead. The distributions are based on computing the delta hedge on the 15th of December, 1995, and evaluating the hedge error in (42) associated with the portfolio on the next trading day, the 18th of December, 1995, and five trading days later, the 22nd of December, 1995. That is,  $\Delta t$  in (42) equals  $\Delta t = 1/365$  and  $5/365$  respectively. The calculations are performed on the prices of contracts traded in the pre-3.00pm period which are common to both pairs of trading days. In computing the delta for the *BS* model,  $M_1$ , the analytical solution for  $\delta_k$  is used; see Hull (2000, p. 312). For the other models, the derivative in (41) is computed numerically. To improve the accuracy of the numerical differentiation, a control variate is used for these models, based on the difference between the *BS* analytical and numerical derivatives. For each value of  $\theta_k$ , the average hedging error over all common contracts is calculated and the density of the (average) hedging error generated as described above.

The means of the hedging distributions are reported in Table 8, with 95% probability intervals given in parentheses. For the densities which are not symmetric and unimodal, these intervals are only approximately equal to 95% HPD intervals. All figures are expressed in cents. It is clear from the results that the location of the hedging distributions is very similar across models. Only the variability differs across models, with the constant volatility models tending to have the most variable hedging error densities, in particular for one day ahead. The exception to this is the  $M_2$  one day ahead hedging error density, which is very tightly concentrated around its mean value. All models produce negative hedging errors one day out and positive hedging errors of a larger magnitude five days out. The *GARCH* models tend to out-perform the constant volatility models one day out, at least in terms of producing hedging errors of a smaller magnitude. However, there is no clear ranking of the models in terms of the five days ahead hedging errors. Most notably, none of the intervals reported in Table 8 cover zero. This can be interpreted as meaning that all models considered are misspecified when it comes to hedging; see also Bakshi, Cao and Chen (1997), who obtain similar qualitative results. However, whether the observed hedging errors are significant from an economic point of view is unclear. The hedging errors range in magnitude from approximately 13 to 52 cents, whilst from Table 1 it can be seen that the option prices in the out of sample dataset themselves range from an average price of \$20.61 for ATM options to an average price of \$70.38 for ITM options. Viewed in relation to the magnitude of the option prices, these hedging errors do not seem to be substantial.

## 5 Conclusions

This paper has developed a Bayesian approach to the implicit estimation of returns models using option-price data. In contrast to existing classical work, the Bayesian method takes explicit account of both parameter and model uncertainty in option pricing. The paper also represents a significant extension of other Bayesian work on option pricing, with a full set of alternative parametric models for returns estimated and ranked using option-price data. Risk-neutral valuation under nonnormal distributional specifications is implemented in a direct and computationally efficient manner.

The results of applying the methodology to 1995 option price data on the S&P500 index show that no one parametric model is ranked highest according to all criteria. The *GST* model clearly dominates all other models, including the *BS* model, in terms of posterior probability, this result being consistent with the excess kurtosis which is estimated from the option prices. The evidence in favour of option-implied skewness is weaker. However, ignor-



Table 8:  
Hedging Performance of the Different Models (cents): One Day and Five Days Ahead  
Means of Hedging Error Densities and 95% Intervals

	One Day Ahead		Five Days Ahead	
	Mean	95% Interval	Mean	95% Interval
$M_1$	-20.195	(-23.500, -16.500)	51.005	(51.001, 51.023)
$M_2$	-14.108	(-14.110, -14.106)	51.677	(51.640, 51.700)
$M_3$	-16.329	(-17.000, -12.500)	52.001	(51.500, 52.080)
$M_4$	-17.308	(-17.900, -16.500)	51.573	(51.000, 52.300)
$M_5$	-13.682	(-13.720, -13.580)	52.056	(52.080, 52.120)
$M_6$	-14.378	(-14.470, -14.370)	51.983	(51.970, 52.020)
$M_7$	-14.454	(-14.500, -14.430)	51.997	(51.960, 51.980)
$M_8$	-13.924	(-13.930, -13.830)	52.310	(52.270, 52.320)
$M_9$	-14.234	(-14.270, -14.230)	51.116	(51.114, 51.118)
$M_{10}$	-13.501	(-13.499, -13.502)	51.677	(51.610, 51.690)

ing the impact of risk factors on the option-based estimates of the higher order moments, it can be concluded that the option prices have factored in more negative skewness than is evident in the symmetric distribution observed for daily S&P500 returns during 1995. This result is consistent with the idea that, since 1987 in particular, option market participants have factored in a larger probability of negative returns than would be predicted by a normal returns distribution; see, for example, Bates (2000). Overall, the constant volatility models which allow for either excess kurtosis or negative skewness in returns, or both, tend to have the best out-of-sample fit and predictive performance, with the *BS* model being ranked lowest in the constant volatility model set on nearly all fit and prediction criteria. The *GARCH* models are assigned virtually zero posterior probability when ranked against the constant volatility models, as well as having an out-of-sample fit and predictive performance which is usually dominated by that of the constant volatility models, including the *BS* model. This inability of the *GARCH* models to capture the behaviour of S&P500 option prices is somewhat consistent with the poor predictive power reported by Chernov and Ghysels (2000) for *GARCH* option pricing models, as based on an earlier sample period for the same option price series. In terms of hedging, all of the models appear to be equally misspecified, although the magnitudes of the hedging errors, relative to the magnitude of the option prices, are very small.

In summary, option market participants appear to have factored in predictions of leptokurtosis and slight negative skewness when pricing the S&P500 options, a conclusion which is clear both from the estimation and out-of-sample results. Time-varying conditional volatility, however, does not appear to be a marked feature of the data. In terms of posterior probability, the model which features symmetry, leptokurtosis and constant volatility over the life of the option, clearly dominates all other contenders. Note however that with option prices being produced via the interaction of market participants invoking potentially different distributional assumptions, option data may well often produce a more even spread of posterior model probabilities than has been observed for this dataset. In this case, an obvious extension of the methodology outlined in the paper would be to invoke the concept of Bayesian model averaging. In particular, the model-averaged predictive, constructed as a weighted average of the model-specific predictives with the relevant model probabilities as weights, may well serve as a more accurate predictive tool than that associated with any one individual model.

## References

- [1] Backus, D., Foresi, S., Li, K. and Liuren, W. (1997), "Accounting for Biases in Black-Scholes", Draft Paper, Stern School of Business, New York University.
- [2] Bakshi, G., Cao, C. and Chen, Z. (1997), "Empirical Performance of Alternative Option Pricing Models", *Journal of Finance*, 52, 2003-2049.
- [3] Bates, D.S. (2000), "Post-87 Crash Fears in the S&P 500 Futures Option Market", *Journal of Econometrics*, 94, 181-238.
- [4] Bauwens, L. and Lubrano, M. (2002), "Bayesian Option Pricing Using Asymmetric GARCH Models", *Journal of Empirical Finance*, 9, 321-342.
- [5] Black, F. and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637-659.
- [6] Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-328.
- [7] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992), "ARCH Modelling in Finance. A Review of the Theory and Empirical Evidence", *Journal of Econometrics* 52, 5-59.

- [8] Bollerslev, T. and Mikkelsen, H. (1999), “Long-term Equity Anticipation Securities and Stock Market Volatility Dynamics”, *Journal of Econometrics*, 92, 75-99.
- [9] Boyle, P.P. and Ananthanarayanan, A.L. (1977), “The Impact of Variance Estimation in Option Valuation Models”, *Journal of Financial Economics*, 375-387.
- [10] Chernov, M. and Ghysels, E. (2000), “A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation”, *Journal of Financial Economics*, 56, 407-458.
- [11] Cho, Y-H. and Engle, R.F. (1999), “Modeling the Impacts of Market Activity on Bid-ask Spreads in the Option Market”, *NBER Working Paper Series*, No. 7331.
- [12] Corrado, C.J. and Su, T. (1997), “Implied Volatility Skews and Stock Index Skewness and Kurtosis implied by S&P 500 index Option Prices”, *Journal of Derivatives*, Summer, 8-19.
- [13] Duan, J-C. (1995), “The GARCH Option Pricing Model”, *Mathematical Finance*, 5, 13-32.
- [14] Duan, J-C. (1999), “Conditionally Fat-Tailed Distributions and the Volatility Smile in Options”, *Draft Paper*.
- [15] Dumas, B., Fleming, J. and Whaley, R. (1998), “Implied Volatility Functions: Empirical Tests”, *Journal of Finance*, 53, 2059-2106.
- [16] Dutta, K.K. and Babbel, D.F. (2002), “Extracting Probabilistic Information from the Prices of Interest Rate Options: Tests of Distributional Assumptions”, *Working Paper 02-26*, The Wharton School, University of Pennsylvania.
- [17] Engle, R.F. (1982), “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation”, *Econometrica*, 50, 987-1008.
- [18] Engle, R.F. and Mustafa, C. (1992), “Implied ARCH Models From Option Prices”, *Journal of Econometrics*, 52, 289-311.
- [19] Eraker, B. (2001), “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices”, *Draft Paper*, Chicago Business School.
- [20] Fernández, C. and Steel, M.F.J. (1998), “On Bayesian Modelling of Fat Tails and Skewness”, *Journal of the American Statistical Association*, 93, 359-371.

- [21] Forbes, C.S., Martin, G.M. and Wright, J. (2002), “Bayesian Estimation of a Stochastic Volatility Model Using Option and Spot Prices”, Department of Econometrics and Business Statistics Working Paper 2/2002, Monash University.
- [22] George, T.J. and Longstaff, F.A. (1993), “Bid-ask Spread and Trading Activity in the S&P 100 Index Options Market”, *Journal of Financial and Quantitative Analysis*, 28, No. 3, 381-397.
- [23] Geweke, J. (1999), “Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication”, *Econometric Reviews*, 18, 1-73.
- [24] Gouriéroux, C. and Monfort, A. (1994), *Simulation Based Econometric Methods*, CORE Discussion Paper.
- [25] Hafner, C.M. and Herwartz, H. (2001), “Option Pricing Under Linear Autoregressive Dynamics, Heteroskedasticity and Conditional Leptokurtosis”, *Journal of Empirical Finance*, 8, 1-34.
- [26] Heston, S. L. and Nandi, S. (2000), “A Closed-Form GARCH Option Valuation Model”, *Review of Financial Studies*, Fall, 13, 585-625.
- [27] Hull, J.C. (2000), *Options, Futures, and Other Derivative Securities*, 4rd ed., Prentice Hall, New Jersey.
- [28] Hull, J.C. and White, A. (1987), “The Pricing of Options on Assets with Stochastic Volatilities”, *Journal of Finance*, 42, 281-300.
- [29] Jacquier, E. and Jarrow, R. (2000), “Bayesian Analysis of Contingent Claim Model Error”, *Journal of Econometrics*, 94, 145-180.
- [30] Jackwerth, J. C. and Rubenstein, M. (2001), “Recovering Stochastic Processes from Option Prices”, Research Paper, University of Wisconsin.
- [31] Jarrow, R. and Rudd, A. (1982), “Approximate Option Valuation for Arbitrary Stochastic Processes”, *Journal of Financial Economics*, 10, 347-369.
- [32] Jones, C. (2000), “The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets”, Draft Paper, University of Rochester.
- [33] Korolyi, G. A. (1993), “A Bayesian Approach to Modelling Stock Return Volatility for Option Valuation”, *Journal of Financial and Quantitative Analysis*, 579-594.

- [34] Lim, G.C., Lye, J.N., Martin, G.M. and Martin, V.L. (1998), “The Distribution of Exchange Rate Returns and the Pricing of Currency Options”, *Journal of International Economics*, 45, 351-368.
- [35] Lim, G.C., Martin, G.M. and Martin, V.L. (2002a), “Parametric Pricing of Higher Order Moments in S&P500 Options”, Department of Econometrics and Business Statistics Working Paper, 1/2002, Monash University.
- [36] Lim, G.C., Martin, G.M. and Martin, V.L. (2002b), “Pricing Currency Options in Tranquil Market: Modelling Volatility Frowns”, Department of Econometrics and Business Statistics Working Paper, 4/2002, Monash University.
- [37] Lye, J.N., Martin, V.L. (1993), “Robust Estimation, Nonnormalities and Generalized Exponential Distributions”, *Journal of the American Statistical Association*, 88, 253-259.
- [38] Lye, J.N., Martin, V.L. (1994), “Nonlinear Time Series Modelling and Distributional Flexibility”, *Journal of Time Series Analysis*, 88, 253-259.
- [39] Pan, J. (2002), “The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-series Study”, *Journal Of Financial Economics* 63, 3-50.
- [40] Polson, N. G. and Stroud, J. R. (2002), “Bayesian Inference for Derivative Prices”, To appear, *Bayesian Statistics* 7.
- [41] Pastorello, S., Renault, E. and Touzi, N. (2000), “Statistical Inference for Random-Variance Option Pricing”, *Journal of Business & Economic Statistics*, 18, 358-367.
- [42] Sabbatini, M. and Linton, O. (1998), “A GARCH Model of the Implied Volatility of the Swiss Market Index from Option Prices”, *International Journal of Forecasting*, 14, 199-213.