Implicit Bayesian Inference Using Option Prices

Gael M. Martin,* Catherine S. Forbes† and Vance L. Martin‡

June 3, 2004

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 predictive and hedging performance. The methodology accom-
modates 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
that accommodate leptokurtosis and skewness 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.

*Department of Econometrics and Business Statistics, Monash University. Corresponding author: Depart-
ment of Econometrics and Business Statistics, P.O. Box, 11E, Monash University, Victoria, 3800, Australia.
(Email: Gael.Martin@Buseco.monash.edu.au).
†Department of Econometrics and Business Statistics, Monash University.
‡Department of Economics, University of Melbourne.
1 Introduction

An option is a derivative asset that gives the owner the right either to buy or sell an underlying asset at some future point in time, at a pre-specified price. The value of the option is thus dependent on the expected future value of the underlying asset. As such, observed market option prices contain information on the underlying price process that is potentially different from and more complete than, information contained in an historical time series of asset prices; see, for example, Pastorello, Renault and Touzi (2000). In this paper, a methodology is presented for conducting inference about a range of models for the underlying price process or, more specifically, for the underlying returns process, using option price data. As is conventional, the inference is referred to as ‘implicit’ in order to contrast it with the more ‘direct’ form of inference based on historical returns 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 adopted as the basis for defining the pertinent risk-neutral process in the estimation of all models that 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.

Predictive densities are produced to assess the out-of-sample performance of the different parametric models. In addition, the hedging performance of the different models is 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 predictive and hedging error assessments are based on data recorded during the week immediately succeeding the end of the estimation period.

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) and Pan (2002) use more formal classical methods to produce implicit estimates of the parameters of stochastic volatility models, with random jumps also accommodated in the case of Bates and Pan. In Lim et al (1998), Bollerslev and Mikkelsen (1999), Duan (1999) and Hafner and Herwartz (2001), \textit{GARCH} models are augmented with nonnormal conditional errors and the implications of such models for option pricing investigated, again within a classical inferential framework. Estimation of the unknown parameters, however, is based on historical returns data only. In Corrado and Su (1997), Dutta and Babbel (2002) and Lim, Martin and Martin (2004), option price data are used to conduct classical implicit estimation of returns models that accommodate skewness and leptokurtosis, with the time-varying volatility formulation of Rosenberg and Engle (1997) and Rosenberg (1998) also specified in the case of Lim, Martin and Martin.

Some Bayesian analyses of option price models 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 (\textit{BS}) model. Bauwens and Lubrano (2002) also use returns data to conduct Bayesian inference, but allow for deviations from the \textit{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 \textit{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 (2003), Polson and Stroud (2002) and Eraker (2003) use both spot and option prices to estimate stochastic volatility models, applying Bayesian simulation methods, with Eraker allowing for random jumps in both returns and volatility. In contrast with the computational complexity associated with these analyses, the methodology outlined in this paper allows for the production of Bayesian posterior quantities using simple low-dimensional deterministic integration techniques.

The paper is organized as follows. Section 2 defines an option price and makes clear the sense in which the price is dependent upon the assumed process for the underlying asset price and the associated unknown parameters. Alternative specifications for returns on the underlying asset that allow for time-varying volatility and nonnormality in the conditional distribution, are formulated in Section 3, with the appropriate risk-neutral adjustments detailed in the Appendix. In Section 4, the manner in which observed market option prices
are used to produce a posterior density function for the parameters of the returns process
is outlined. The methodology is then illustrated in Section 5 using option price data on the
S&P500 index. Posterior quantities are reported, together with summary measures of the
predictive and hedging distributions for the different models. The empirical results provide
evidence that option prices have factored in the assumption of a returns distribution with
excess kurtosis and a small amount of negative skewness. Little support is found for the
GARCH specifications. The hedging results suggest that the hedging errors for all models
are insubstantial. Some conclusions are drawn in Section 6.

2 Option Pricing

An option is an asset that gives the owner the right to either buy (call option) or sell (put
option) the underlying asset at some future point in time, at some pre-specified exercise, or
strike price, $K$. For the European call option that is the focus of this paper, the owner has
the right to buy the underlying asset (or exercise the option) only when the option matures,
at the future time point $T$. In this case, the price of the option is given by the expected value
of the discounted payoff of the option,

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

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, Chp. 11). The notation used
in (1) is defined as follows:

\begin{align*}
T & \quad \text{the time at which the option is to be exercised;} \\
\tau & \quad \text{the length of the option contract;} \\
K & \quad \text{the exercise, or strike price at which the underlying asset is to be bought;} \\
S_T & \quad \text{the spot price of the underlying (non-dividend paying) asset at the time of maturity;} \\
r & \quad \text{the risk-free interest rate assumed to hold over the life of the option.}
\end{align*}

The option price is thus a function of the observable quantities, $r$, $K$ 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 the spot price; and (ii)
the values assigned to the unknown parameters of that 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 for the spot price and posterior probability distributions for the model specific parameters.

The dependence of the option price $q$ on the parameters that characterize the underlying spot price process is made explicit by re-writing (1) as

$$q(\theta) = e^{-r\tau} \int_{K}^{\infty} (S_T - K)p(S_T|S_t; \theta)dS_T,$$

where $p(S_T|S_t; \theta)$ is the risk-neutral probability density function (pdf) of the spot price at the time of maturity of the option, conditional on the current price $S_t$, with $\theta$ being the vector of unknown parameters that feature in that pdf. For example, in the case of the BS pricing model, in which the spot price is assumed to follow a geometric Brownian motion process with diffusion (or volatility) parameter $\sigma$, $p(S_T|S_t; \theta = \sigma)$ is lognormal and the integral in (2) has the solution,

$$q(\sigma) = S_t \Phi(d_1) - K e^{-r\tau} \Phi(d_2),$$

where $d_1 = (\ln(S_t/K) + (r + 0.5\sigma^2) \tau)/\sigma \sqrt{\tau}$, $d_2 = d_1 - \sigma \sqrt{\tau}$ and $\Phi(.)$ denotes the cumulative normal distribution function. In the more general distributional settings to be entertained in this paper, neither $p(S_T|S_t; \theta)$ nor $q(\theta)$ can be expressed in closed-form, with $q(\theta)$ to be evaluated instead via simulation in a manner to be described in Section 5.

3 Distributional Specifications for Returns

The geometric Brownian motion assumption underlying the BS model implies that in discrete time the continuously compounded return, defined as the log-differenced spot price, is normally distributed with constant volatility. 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 implied volatilities are not constant across strike prices or maturity. That is, when the BS price in (3) is equated with an observed option price, the implied estimate of $\sigma$ is found to vary across different values of the strike price $K$. These implied volatility ‘smiles’ or ‘smirks’, which have also been found to vary in intensity depending on the time to expiration $\tau$, have become a stylized fact in empirical work on option prices. Such patterns have been shown to be evidence of implied returns models that deviate from the specifications of the BS model; see, for example, Bakshi, Cao and Chen
In this section the assumptions that 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 that 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. Adoption of more general volatility models, such as those used in the returns-based analyses of Bollerslev and Mikkelsen (1999) and Bauwens and Lubrano (2001), has computational implications for the option price-based estimation method advocated in this paper. Hence, the GARCH(1,1) model is retained as a parsimonious omnibus model of volatility.

Under the assumption that no riskless arbitrage opportunities exist, a portfolio comprised of both the option and the underlying asset can be constructed, such that the return on the portfolio is equal to the risk-free rate of interest, \( r \). The existence of such a portfolio implies that options are priced as if investors were risk-neutral, with the expectation in (1) evaluated not under the empirical, or ‘objective’ distribution, but under the risk-neutral distribution for the spot price as a consequence; see Campbell, Lo and McKinley (1997, Chp. 9) and Hull (2000, Chp. 11) for details. In the case of the BS model, risk-neutral pricing is effected by shifting the drift term in the continuous time diffusion process for the logarithm of the spot price by a factor of \( r - \mu \), where \( \mu \) is the actual rate of return on the underlying asset. For the more general pricing framework entertained here, a modified version of the approach of Duan (1995, 1999) is applied in order to derive the appropriate risk-neutral distribution. Details of this derivation are provided in the Appendix. As is the case for the BS model (which is nested in the general specification), the move from the empirical to risk-neutral models involves a mean shift equivalent to the difference between the risk-free and actual rates of return on the underlying asset.

The discrete time version of the risk-neutral distribution that underlies the BS option price is given by

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

where \( \ln S_{t+\Delta t} - \ln S_t \) denotes the continuously compounded return over the small time interval \( \Delta t \), \( r_{t+\Delta t} \) is the risk-free rate of return at time \( t + \Delta t \), \( z_{t+\Delta t} \) is an iid \( N(0,1) \) variate and \( \sigma \) is the volatility of returns. By convention, both \( r_{t+\Delta t} \) and \( \sigma \) are treated as annualized quantities. The model adopted in this paper generalizes (4) to
\[ \ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5\sigma_{t+\Delta t}^2)\Delta t + \sigma_{t+\Delta t} \sqrt{\Delta t} v_{t+\Delta t}, \]  

where the time-varying conditional variance, \( \sigma_{t+\Delta t}^2 \) is assumed to follow a GARCH(1,1) process,

\[ \sigma_{t+\Delta t}^2 = \frac{\alpha_0}{\Delta t} + \alpha_1 \sigma_t^2 v_t^{(a)} + \alpha_2 \sigma_t^2, \]  

with \( \alpha_0 > 0; \alpha_1, \alpha_2 \geq 0 \) and \( \alpha_1 + \alpha_2 < 1 \). The innovation term \( v_{t+\Delta t} \) in (5) is an iid standardized nonnormal variate to be specified below and the term \( v_t^{(a)} \) in (6) denotes the appropriately risk-adjusted innovation, \( v_t^{(a)} = v_t - \lambda v_t^N \), where \( \lambda v_t^N = \sqrt{\Delta t} (\mu_t - r_t) / \sigma_t \) and \( \mu_t \) is the conditional mean of the return; see the Appendix for details. By definition, the parameters of the distribution of \( v_t \), 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.

Denoting by \( w_t \) a random variable with mean \( \mu_w \) and variance \( \sigma_w^2 \), and defining the standardized variate \( v_t \) via \( w_t = \sigma_w v_t + \mu_w \), the approach of Fernandez and Steel (1998) is used to define the pdf of \( v_t \) as

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

where \( f(.) \) is defined as a symmetric pdf with a single mode at zero and \( I_A(w_t) \) denotes the indicator function for the set \( A \). The mean and variance of \( w_t \) are defined respectively as \( \mu_w = \left( \frac{\gamma^2 - 1/\gamma}{\gamma + 1/\gamma} \right) \int_0^\infty 2x f(x) dx \) and \( \sigma_w^2 = \left( \frac{\gamma^3 + 1/\gamma^3}{\gamma + 1/\gamma} \right) \left( \int_0^\infty 2x^2 f(x) dx \right) - \mu_w^2 \). 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 pdf \( p_f(v_t) \) has a mean of zero, with the sign and magnitude of \( \gamma - 1 \) determining the sign and magnitude of the mode.

The pdf in (7) can be used to produce a standardized skewed normal distribution for \( v_t \) when \( f(.) \) is the normal density function. Alternatively, defining \( f(.) \) as a pdf 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. For a leptokurtic specification for \( f(.) \) 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; see Duan (1999) on this point. Defining a random variable \( \eta_t \) with mean and variance \( \mu_\eta \) and \( \sigma_\eta^2 \) respectively, this pdf is given by

\[ f(\eta_t) = k^* (1 + \frac{\eta_t^2}{\nu})^{-0.5(\nu+1)/2} \exp(-0.5\eta_t^2), \]  

7
where \( k^* = \left[ f(1 + \frac{\eta^2}{2})^{-0.5(\nu+1)/2} \exp(-0.5\eta^2) d\eta \right]^{-1} \) is the normalizing constant. The density in (8) 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 (8) also exist for any value of \( \nu \). We refer to the pdf in (8) as the Generalized Student \( t \) (GST) density. In order to define a GST density for the standardized variate \( \eta_t \), defined by \( \eta_t = \sigma_{\eta} v_t + \mu_{\eta} \), the variance of \( \eta_t \), \( \sigma^2_{\eta} \), needs to be computed numerically, along with the integrating constant \( k^* \) in (8). The mean of \( \eta_t \), \( \mu_{\eta} \), is equal to zero.

4 Bayesian Inference Using Observed Option Prices

The general distributional framework outlined in the previous section nests several alternative specifications for returns. The precise set of models that are entertained for the particular data set under consideration are detailed in Section 5.1 below. At this point we simply emphasize that each model for returns implies a particular specification for \( q(\theta) \) in (2), with attendant parameter vector \( \theta \). Posterior inferences about \( \theta \) are to be produced implicitly from observed market option prices, via a model that describes how observed prices deviate from the theoretical price \( q(\theta) \). 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. Moneyness is defined in terms of the ratio of the current spot price, \( S_t \), to the strike price of the option, \( K \), where \( S_t - K \) is a measure of the payoff of the option if it were to be exercised at time \( t \). Broadly speaking, options are said to be out-of-the-money (OTM) if \( S_t < K \), at-the-money (ATM) if \( S_t = K \) and in-the-money (ITM) if \( S_t > K \). 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 category \( j \), \( j = 1, 2, \ldots, J \), 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 categories and the location of segment boundaries, \( m_j, j = 0, 1, 2, \ldots J \), are chosen to accord with the main moneyness groups in the data. More details of this are provided in Section 5. 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 $t = 1, 2, \ldots, n$, and $n$ is the number of trading days used in the estimation sample.

The number of observations in each moneyness category at each point in time, $n_{jt}$, varies. Letting $i = 1, 2, \ldots, n_{jt}$, $j = 1, 2, \ldots, J$, $t = 1, 2, \ldots, n$, the total number of observations in the sample is given by $N = \sum_{j=1}^{J} \sum_{t=1}^{n} n_{jt}$. The model specified for the $N$ observed option prices is

$$C_{ijt} = b_{0j} + b_{1j}q(z_{ijt}, \theta) + \sum_{l=1}^{4} d_{lj}D_{lt} + \sum_{g=1}^{G} \rho_{gj}C_{ij(t-g)} + \psi_{j}u_{ijt},$$  \hspace{1cm} (9)$$

$$u_{ijt} \sim iid \ N(0, 1) \text{ for all } i = 1, 2, \ldots, n_{jt}; \ j = 1, 2, \ldots, J; \ t = 1, 2, \ldots, n. \hspace{1cm} (10)$$

The function $q(.,.)$ in (9) represents the theoretical option price evaluated as per (2), where it is now made explicit that $q(.,.)$ depends upon both the parameter vector $\theta$ that characterizes the assumed returns model, and the vector of observables, $z_{ijt} = (r_{t}, K_{ij}, \tau_{ij}, S_{t})'$, 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 (9) allows an observed option price to deviate from the theoretical price in a manner that differs across moneyness category. In particular, allowance for heteroscedasticity across moneyness categories 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, $D_{lt}$, $l = 1, 2, 3, 4$, where Friday corresponds to $D_{lt} = 0$ 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 category $j$, $g = 1, 2, \ldots, G$, 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 may be grouped together by moneyness category, and denoted by $\beta_{j} = (b_{0j}, b_{1j}, d_{1j}, d_{2j}, d_{3j}, d_{4j}, \rho_{11j}, \ldots, \rho_{Gj})'$, for $j = 1, \ldots, J$, with $\beta = (\beta_{1}', \beta_{2}', \ldots, \beta_{J}')'$. The variances associated with each moneyness category may also be grouped as $\Sigma = diag(\psi_{1}^{2}, \ldots, \psi_{J}^{2})$. Further defining $c_{j}$ as the $(N_{j} \times 1)$ vector of observed options prices for category $j$, ordered by day within the category, with $N_{j} = \sum_{t=1}^{n} n_{jt}$, given the distributional assumption in (10) the likelihood function is given by
\[
L(\Sigma, \beta, \theta) = (2\pi)^{-N/2} \prod_{j=1}^{J} \psi_j^{-N_j} \exp \left(-\frac{1}{2\psi_j^2} \left[ c_j - X_j(\theta)\beta_j \right]' \left[ c_j - X_j(\theta)\beta_j \right] \right), \quad (11)
\]

where \(X_j(\theta)\) is an \((N_j \times L)\) matrix containing the values of the \(L = 6 + G\) regressors in (9), for category \(j\), again ordered by day within the group, and the likelihood is conditional on initial values for the lagged option prices. For notational convenience we do not make explicit the dependence of \(L(.)\) on the values of all observable components on the right hand side of (9). We also omit these components in the subsequent description of all posterior densities.

Assuming a joint prior pdf for \(\beta\) and \(\Sigma\) of the form

\[
p(\beta, \Sigma) \propto \prod_{j=1}^{J} \psi_j^{-2}, \quad (12)
\]

and imposing a priori independence between \((\beta, \Sigma)\) and \(\theta\), standard Bayesian algebra associated with the normal linear regression structure of (9) can be used to integrate out \((\beta, \Sigma)\) from the joint posterior pdf, defined by the product of (11), (12) and the prior pdf on \(\theta\), \(p(\theta)\). This leads to the posterior pdf for \(\theta\),

\[
p(\theta|c) \propto \prod_{j=1}^{J} \left| X_j(\theta)'X_j(\theta) \right|^{-1/2} \psi_j^{-(N_j-L)} \times p(\theta), \quad (13)
\]

where \(c = (c_1', c_2', \ldots, c_J')', \伽 = \left[ c_j - X_j(\theta)\hat{\beta}_j \right]' \left[ c_j - X_j(\theta)\hat{\beta}_j \right] / (N_j - L)\) and \(\hat{\beta}_j = [X_j(\theta)'X_j(\theta)]^{-1} X_j(\theta)'c_j\).

Due to the nonstandard nature of the pdf in (13), numerical procedures are required in order to produce all posterior quantities of interest related to \(\theta\). Those quantities comprise marginal posterior densities for the individual elements of \(\theta\), with associated moments; posterior model probabilities; predictive densities for future option prices; and the posterior densities for the hedging errors. Details of these procedures are provided in Section 5. All calculations are performed using the GAUSS software, with the programs being available from the authors on request.

5 Empirical Application Using S&P500 Option Prices

5.1 Detailed Model Specifications

In this section, S&P500 option price data are used to conduct implicit Bayesian inference on a range of alternative returns models that are nested in the distributional framework outlined in Section 3. Associated with the assumption of constant volatility \((\sigma_t = \sigma)\) in (5) are four alternative models for returns, corresponding to the alternative specifications for \(f(.)\) and \(\gamma\).
in (7): normal, $GST$, skewed normal ($SN$) and skewed $GST$ ($SGST$), denoted respectively by $M_1$, $M_2$, $M_3$ and $M_4$:

\begin{align*}
M_1 & : f(.) \text{ normal; } \gamma = 1; \quad \sigma_t = \sigma \quad v_t \sim N(0,1) \\
M_2 & : f(.)GST; \quad \gamma = 1; \quad \sigma_t = \sigma \quad \mu_\eta + \sigma_\eta v_t \sim GST(\mu_\eta, \sigma^2_\eta, \nu) \\
M_3 & : f(.) \text{ normal; } \gamma \neq 1; \quad \sigma_t = \sigma \quad \mu_w + \sigma_w v_t \sim SN(\mu_w, \sigma^2_w, \gamma) \\
M_4 & : f(.) 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^2_w, \gamma, \nu).
\end{align*}

(14)

As model $M_1$ corresponds to the discrete time version of the returns model that 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 (6) 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, $GST$, $SN$ and $SGST$.

In order to reduce the number of unknown parameters and, hence, the computational demands of the proposed estimation method, certain restrictions are placed on the parameterization of the $GARCH$ models. First, the intercept parameter $\alpha_0$ in (6) 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, six $GARCH$ models are estimated, denoted respectively by $M_5$, $M_6$, $M_7$, $M_8$, $M_9$ and $M_{10}$:

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

(15)

Model $M_5$ specifies a conditional normal distribution, whilst for models $M_6$ to $M_{10}$ nonnormal conditional distributions are specified. The models that accommodate excess kurtosis in the conditional distribution of $v_t$ ($M_6$, $M_7$, $M_9$ and $M_{10}$) are estimated with $\nu$ set to either 1.0 or 5.0. The values of $\nu$ are chosen so as to produce a continuum of kurtosis behaviour in the conditional distribution, with kurtosis measured by a numerical estimate of $E(v_t^4)$, with $v_t$ as defined in (5). For $\nu = 5.0$, kurtosis is equal to 3.233, whilst for $\nu = 1.0$, it is equal to 3.624, both values of $\nu$ implying kurtosis in excess of the value of 3 associated with the normal distribution. 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 captures some of the kurtosis in the unconditional distribution. Since the GARCH model does not accommodate asymmetry in returns, it is legitimate to specify a degree of skewness in the associated conditional distribution that is equivalent to that in the unconditional distribution of the corresponding constant volatility model. Hence, the model that specifies GARCH with a conditional SN distribution ($M_8$) is estimated with $\gamma$ set to the value estimated for the corresponding model with constant volatility ($M_3$). This value of $\gamma = 0.85$ corresponds to a skewness coefficient of $-0.253$, where the latter is an estimate of $E(v^3_t)$. The degree of skewness specified for the GARCH models with the SGST conditional distributions ($M_9$ and $M_{10}$) also matches that estimated for the corresponding constant volatility model ($M_4$), with the value of $\gamma = 0.80$ corresponding to a skewness coefficient of $-0.341$.

For the BS model, $M_1$, the analytical expression for $q(z_{ijt}, \theta)$ in (9) is as given in (3), with $\theta = \sigma$. For all other models, $q(z_{ijt}, \theta)$ does not have a closed-form solution and, hence, needs to be estimated via simulation. For models $M_2$ to $M_4$ the approach adopted is to simulate (5) over the life of the contract, with $\sigma_t = \sigma$, and with the innovations drawn from the relevant nonnormal distribution in (14). For each of these models, simulation of the relevant process for returns is repeated $h$ times, producing $S^{(l)}_T, l = 1, 2, \ldots, h$, and the expectation in (1) approximated by the sample mean of $e^{-r_t \tau_{ij}} \max(S^{(l)}_T - K_{ij}, 0)$. 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 $M_{10}$, both processes in (5) and (6) are simulated over the life of the option. Since a constant mean is found to be appropriate for the S&P500 data set under consideration, $\mu_t$ in the risk-adjusted innovation in (6) is replaced by the sample mean of returns on the index for 1995. For a general discussion of this simulation-based approach to the pricing of options see Gourieroux and Monfort (1994). 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, $\Delta t$ equates with 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.
5.2 Data Description

The data are based on bid-ask quotes on European 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 the few minutes immediately prior to 3.00pm on each day. As noted earlier, this form of data selection is aimed at minimizing the amount of intraday variation in the spot prices recorded synchronously with the option prices. A cross section of approximately 40 prices is selected on each day, with the prices deliberately chosen so as to span the full moneyness spectrum. Following Bakshi, Cao and Chen (1997), options for which \( S_t/K_{ij} \in (0.98, 1.04) \) are categorized as ATM, those for which \( S_t/K_{ij} \leq 0.98 \), as OTM, and those for which \( S_t/K_{ij} \geq 1.04 \), as ITM. 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 data set 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, \( S_{te} \), where \( D = 0.026 \) is the average annualized dividend rate paid over the life of the option, 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

\[
l_{bij} = \max\{0, S_{te}^{-D\tau_{ij}} - e^{r\tau_{ij}}K_{ij}\}
\]

leaves \( N = 8968 \) 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 applied to the estimation data set. This yields a total of 984 option prices to be used in assessing the out-of-sample performance of the models. The characteristics of this data set 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 166 out-of-sample options traded tend toward the higher end of that range, with the average price and bid-ask spread being larger as a consequence, than the corresponding figures in the
Table 1: S&P500 Option Price Dataset

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

\(^{(a)}\) Not applicable.

estimation sample. There are 818 ITM options traded in the out-of-sample period, and their average prices and bid-ask spreads are very similar to those of the estimation period.

### 5.3 Priors for the Elements of \(\theta\)

The 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. 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. However, comparison of the constant volatility models with any of the time varying volatility models is potentially affected by the choice of finite integration bounds for \(\sigma\) used in the construction of the Bayes factors (to be described in Section 5.5.1 below). Investigation of this potential impact using the criterion of Kass (1993) leads to the conclusion that all Bayes factors and associated model probabilities reported below are robust to the choice of bounds for \(\sigma\).
An inverted gamma prior is specified for $\nu$, with $E(\nu) = 1.76$ and $\text{var}(\nu) = 197.89$. 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 $\text{var}(\gamma) = 1.0$. 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. For the $GARCH$ models, a uniform prior is placed on the joint space of $\alpha_1$ and $\alpha_2$, bounded by $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$. Since these bounds on $\alpha_1$ and $\alpha_2$ have a well-defined interpretation in the context of the $GARCH$ model, the specification of a joint uniform prior over this region does not introduce any element of arbitrariness into the construction of the Bayes factors.

### 5.4 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 $M_k$, $k = 1, 2, \ldots, 10$, the joint posterior pdf for $\theta_k$, $p(\theta_k|c)$, is given by (13), with $c$ denoting the vector of 8968 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. 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 $M_{10}$. The advantage of this deterministic approach is that the results produced are essentially exact, with none of the convergence issues that would be associated with a Markov chain Monte Carlo (MCMC) 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, as discussed in Section 5.1 above. That is, it would not be computationally feasible to produce the number of Markov chain iterates required to establish convergence, in combination with the simulation-based estimation of the theoretical option prices.\(^3\)

In Table 2, the mean, mode and approximate 95% Highest Posterior Density (HPD) intervals are reported for each parameter in the ten models estimated. 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 that is as close to the nominal coverage as possible given the discrete grid defined for each parameter.
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 of 3.674, 3.645 and 3.650 respectively. Remembering that, by construction, both $\nu$ and $\gamma$ are interpreted as distributional parameters for implicit (risk-neutral) 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 $\nu$ that translate into kurtosis values that 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 skewness coefficients 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 $\gamma$ that imply either symmetry ($\gamma = 1$) or positive skewness ($\gamma > 1$), in addition to values implying negative skewness ($\gamma < 1$).

For all six time-varying volatility models, $M_5$ to $M_{10}$, the option-implied persistence in daily volatility are relatively low, with point estimates, given by $\hat{\alpha}_1 + \hat{\alpha}_2$, ranging in value from 0.8 to 0.84. Similar values (not reported) are obtained from a GARCH(1,1) analysis of S&P500 index daily returns over the period 1994 to 1997. The small values estimated for $\alpha_1$ indicate that the implied volatility process for the S&P500 index evolves relatively smoothly over the life of the option. By construction, the long-run volatility is held fixed at an annualized value of 0.12 in all cases.

5.5 Model Rankings

5.5.1 Implicit Model Probabilities

Implicit model probabilities are derived from the posterior odds ratios, constructed for each model, $M_2$, $M_3$, ..., $M_{10}$, relative to a reference model, $M_1$. Defining $P(M_k|c)$ as the posterior probability of $M_k$, the posterior odds ratio for $M_k$ versus $M_1$ is given by

$$\frac{P(M_k|c)}{P(M_1|c)} = \frac{P(M_k)}{P(M_1)} \times \frac{p(c|M_k)}{p(c|M_1)} = \text{Prior Odds} \times \text{Bayes Factor}$$

(17)
Table 2:
Implicit Marginal Posterior Densities\textsuperscript{(a)}

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Mode</th>
<th>Mean</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>$\sigma$</td>
<td>0.115</td>
<td>0.115</td>
<td>(0.106, 0.124)</td>
</tr>
<tr>
<td>$M_2$</td>
<td>$\sigma$</td>
<td>0.125</td>
<td>0.125</td>
<td>(0.116, 0.134)</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>0.850</td>
<td>0.934</td>
<td>(0.450, 1.650)</td>
</tr>
<tr>
<td>$M_3$</td>
<td>$\sigma$</td>
<td>0.115</td>
<td>0.115</td>
<td>(0.106, 0.124)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>0.850</td>
<td>0.986</td>
<td>(0.400, 1.600)</td>
</tr>
<tr>
<td>$M_4$</td>
<td>$\sigma$</td>
<td>0.115</td>
<td>0.115</td>
<td>(0.106, 0.124)</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
<td>0.850</td>
<td>0.919</td>
<td>(0.250, 2.100)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>0.800</td>
<td>0.891</td>
<td>(0.650, 1.150)</td>
</tr>
<tr>
<td>$M_5$</td>
<td>$\alpha_1$</td>
<td>0.030</td>
<td>0.031</td>
<td>(0.022, 0.038)</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2$</td>
<td>0.810</td>
<td>0.810</td>
<td>(0.802, 0.818)</td>
</tr>
<tr>
<td>$M_6$</td>
<td>$v = 5.0$</td>
<td>$\alpha_1$</td>
<td>0.030</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2$</td>
<td>0.810</td>
<td>0.810</td>
</tr>
<tr>
<td>$M_7$</td>
<td>$v = 1.0$</td>
<td>$\alpha_1$</td>
<td>0.031</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2$</td>
<td>0.810</td>
<td>0.810</td>
</tr>
<tr>
<td>$M_8$</td>
<td>$\gamma = 0.85$</td>
<td>$\alpha_1$</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2$</td>
<td>0.760</td>
<td>0.076</td>
</tr>
<tr>
<td>$M_9$</td>
<td>$v = 5.0; \gamma = 0.80$</td>
<td>$\alpha_1$</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2$</td>
<td>0.780</td>
<td>0.780</td>
</tr>
<tr>
<td>$M_{10}$</td>
<td>$v = 1.0; \gamma = 0.80$</td>
<td>$\alpha_1$</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\alpha_2$</td>
<td>0.780</td>
<td>0.780</td>
</tr>
</tbody>
</table>

\textsuperscript{(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, $\alpha_1$ and $\alpha_2$, measures daily persistence in volatility.
for \( k = 2, 3, \ldots, 10 \), where

\[
p(c|M_k) = \int_{\Sigma} \int_{\beta} \int_{\theta_k} L(\Sigma, \beta, \theta_k|M_k)p(\Sigma, \beta, \theta_k|M_k)d\Sigma d\beta d\theta_k,
\]  

(18)

is the marginal likelihood of \( M_k \), with \( L(\Sigma, \beta, \theta_k|M_k) \) and \( p(\Sigma, \beta, \theta_k|M_k) \) denoting respectively the likelihood and prior under \( M_k \). The likelihood is as defined in (11) and the prior is defined by product of (12) and the model-specific \( p(\theta_k) \). The model probabilities are calculated by solving the nine ratios in (17) subject to the normalization \( \sum_{k=1}^{10} P(M_k|c) = 1 \). The models are then ranked as \textit{a posteriori} most probable to least probable according to the size of these probabilities. As \( \Sigma \) and \( \beta \) can be integrated out analytically, the marginal likelihood for model \( M_k \) in (18) reduces to

\[
p(c|M_k) = h \int_{\theta_k} L^*(\theta_k|M_k)p(\theta_k|M_k)d\theta_k,
\]  

(19)

where \( L^*(\theta_k|M_k) = \prod_{j=1}^{J} |X_j(\theta_k)'X_j(\theta_k)|^{-1/2} \psi_j^{-(N_j-L)} \) and \( h \) is a constant that is independent of the specification of \( M_k \). The integral in (19) is that which is computed in the numerical normalization of the posterior density for \( \theta_k \) in (13). 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 in a straightforward manner.

Table 3 provides the estimated Bayes factors for the ten models \( M_1 \) to \( M_{10} \). The final row gives the associated model probabilities for all ten models, based on equal prior probabilities, \( P(M_k) = P(M_1) \), \( k = 2, 3, \ldots, 10 \), in (17). There are three notable aspects of the results in Table 3. First, the symmetric GST model with constant volatility (\( M_2 \)) is assigned all posterior probability (to two decimal places) in the set of ten alternative models, as is evident from the last row in Table 3. This is completely consistent with the fact that the option prices have factored in distributional estimates that 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 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 \( \alpha_1 \) in the GARCH specification in (6), which are, in turn, associated with a smoothly evolving volatility process over the life of
Table 3:
Implicit Bayes Factors and Model Probabilities.
Entry \((i, j)\) Indicates the Bayes Factor in Favour of \(M_j\) Versus \(M_i\)

<table>
<thead>
<tr>
<th>(M_1)</th>
<th>(M_2)</th>
<th>(M_3)</th>
<th>(M_4)</th>
<th>(M_5)</th>
<th>(M_6)</th>
<th>(M_7)</th>
<th>(M_8)</th>
<th>(M_9)</th>
<th>(M_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>31400</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>1200</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>2050</td>
<td>8.3E07</td>
<td>31.30</td>
<td>8.8E09</td>
<td>9.2E25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>40260</td>
<td>0.00</td>
<td>4.3E06</td>
<td>4.5E22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>106</td>
<td>1.1E18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>2.8E08</td>
<td>3.0E24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.1E16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

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 that 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 (\(M_5\)).

5.5.2 Out-of-Sample Predictive Performance

For model \(M_k\), the predictive pdf for the \(i\)th option price \(C_{ijf}\), for moneyness category \(j\), observed on some day \(f\) during the hold-out sample period is given by

\[
p(C_{ijf}|c) = \int_{\beta_j} \int_{\psi_j} \int_{\theta_k} p(C_{ijf}|\beta_j, \psi_j, \theta_k, c)p(\beta_j, \psi_j|\theta_k, c)p(\theta_k|c)d\beta_jd\psi_jd\theta_k,
\]

where \(p(C_{ijf}|\beta_j, \psi_j, c, \theta_k)\) is a normal density, given the assumption of a normal distribution for \(u_{ijf}\) in (10). Standard Bayesian distribution theory for a normal linear model yields a
multivariate Student t posterior distribution for $\beta_j$, conditional on $\theta_k$, with $E(\beta_j|\theta_k, c) = \hat{\beta}_j$ and $\text{var}(\beta_j|\theta_k, c) = \hat{\psi}_j^2 \left[ X_j(\theta_k)' X_j(\theta_k) \right]^{-1}$, where $\hat{\beta}_j$ and $\hat{\psi}_j^2$ are as defined after (13). This result, conditional upon $\theta_k$, implies that the predictive pdf can be expressed as a mixture,

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

where $p(C_{ijf}|\theta_k, c)$ is a univariate Student t density with $E(C_{ijf}|\theta_k, c) = x_{ijf}(\theta_k)' \hat{\beta}_j$ and $\text{var}(C_{ijf}|\theta_k, c) = \hat{\psi}_j^2[1 + x_{ijf}(\theta_k)'(X_j(\theta_k)'X_j(\theta_k))^{-1}x_{ijf}(\theta_k)]$, with $x_{ijf}(\theta_k)'$ denoting the $(1 \times L)$ vector of observations at time period $f$ on the $L = 6 + G$ regressors associated with $C_{ijf}$. Hence, the predictive density in (20) 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 in (16) is imposed before averaging over the space of $\theta_k$.

The estimated predictive pdf 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. The calculation is repeated for all option contracts, the predictive probability recorded for model $M_k$ being the average of all computed probabilities.4 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 that fall within the observed bid-ask spreads. The same calculation is performed for the predictive means. Third, the proportion of market prices that fall within the 95% probability interval associated with the estimated predictive, is calculated for each model. All calculations are performed for the 166 ATM, and 818 ITM contracts, as well as for all 984 contracts in the hold-out sample. However, in Table 4, only results for all contracts are reported, since the model rankings for the ATM and ITM contracts are equivalent.

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 according to all criteria considered. The average probability ascribed to the bid-ask spread ranges from 33.7% for $M_3$ and $M_4$ to 33.9% for $M_2$, with $M_7$ being the next best performer with a coverage of 32.3%. There is little to choose between the GARCH models and the BS model according to this criterion. In terms of the proportion of times that the point predictors, the predictive mean and mode, fall in the bid-ask spread, BS model is the next best performer after models $M_2$ to $M_4$, 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.

For all models, 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
Table 4:
Predictive Performance of the Different Models

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

\(^{(a)}\) $BA$ = the bid-ask spread.
\(^{(b)}\) The 95% Interval is the interval that excludes 2.5% in the lower and upper tails of the predictive distribution. This interval equals the 95% HPD interval only for those predictives that are symmetric around a single mode.

$M_4$. The 95% predictive interval covers the observed market price approximately 70% of the time for all models. Note that the average width of the intervals exceeds the average width of the bid-ask spread by approximately 80%.

5.5.3 Hedging Performance

As noted in Section 3, option pricing is underpinned by the concept of a risk-free portfolio comprised of the option and the underlying asset. In other words, options can provide a hedge against movements in the underlying asset. The relative proportions of the two assets (option and underlying) that are required to render the portfolio risk-free are model dependent. Hence, another measure of the performance of a particular model is the extent to which the return on the hedge portfolio, constructed on the assumption of that model, differs from the risk-free rate of return. Equivalently, the alternative models can be compared in terms of the extent to which their associated hedging errors deviate from zero.

In this paper attention is restricted to delta hedges. The delta for the $ith$ option price, in moneyness group $j$, observed at time $t$, based on the assumption that $M_k$ describes the
returns process, is defined as
\[ \delta_k = \frac{\partial q(z_{ijt}, \theta_k)}{\partial S_t}. \] (21)

In computing the hedging errors, the portfolio consists of going short in (selling) the option and long in (buying) the underlying asset by an amount of \( \delta_k \), 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
\[ H_k = \delta_k \left[ S_{t+\Delta t} - S_t e^{r\Delta t} \right] - \left[ C_{ij(t+\Delta t)} - C_{ijt} e^{r\Delta t} \right]. \] (22)

When the assumptions of the BS model (\( M_1 \)) are valid and when there is continuous re-balancing (\( \Delta t \to 0 \)) there are no hedging errors. When rebalancing is discrete and the BS assumptions are not valid, \( H_k \) will not necessarily be equal to zero. However, it is expected that \( H_k \) will be closer to zero, the better is any particular model at explaining the characteristics of the data; see also Bakshi, Cao and Chen (1997). The posterior distribution of the hedging error in (22) 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 pdf, \( p(H_k|c) \), 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, judged to be 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 (22) associated with the portfolio on the next trading day, the 18th of December, 1995, and five trading days later, the 22nd of December, 1995. The calculations are performed on the prices of contracts traded in the pre-3.00pm period that are common to both pairs of trading days. In computing the delta for the BS model, \( M_1 \), the analytical solution for \( \delta_1 \) is \( \Phi(d_1) \), with the latter as defined following (3); see Hull (2000, p. 312). For the other models, the derivative in (21) 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 5, with 95% probability
intervals given in parentheses. All values are expressed in cents. It is clear from the results that the location of the hedging error distributions is very similar across models, with the exception of the BS model one day ahead where the absolute value of the average hedging error is larger than for all other models. Further, 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 M2 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, again with the exception of M2 whose hedging errors are of a similar magnitude to those of the GARCH models. However, there is no clear ranking of the models in terms of the five days ahead hedging errors. Whilst none of the intervals reported in Table 5 cover zero, it is unclear as to whether this is due to model misspecification, or to a lack of continuous rebalancing. It is presumed that for smaller $\Delta t$, as in the one day ahead case, proportionally more of the hedging error would be due to model misspecification; see also Bakshi, Cao and Chen (1997) for discussion and similar qualitative results. In any case, whether the observed hedging errors are significant from an economic point of view is arguable. 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 data set 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 therefore seem to be substantial.

6 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. The methodology produces all posterior quantities of interest, from marginal parameter densities to hedging error densities, using straight-forward, low-dimensional deterministic integration. 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
Table 5:

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

<table>
<thead>
<tr>
<th></th>
<th>One Day Ahead</th>
<th></th>
<th>Five Days Ahead</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>95% Interval</td>
<td>Mean</td>
<td>95% Interval</td>
</tr>
<tr>
<td>$M_1$</td>
<td>-20.195</td>
<td>(-23.500, -16.500)</td>
<td>51.005</td>
<td>(51.001, 51.023)</td>
</tr>
<tr>
<td>$M_3$</td>
<td>-16.329</td>
<td>(-17.000, -12.500)</td>
<td>52.001</td>
<td>(51.500, 52.080)</td>
</tr>
<tr>
<td>$M_4$</td>
<td>-17.308</td>
<td>(-17.900, -16.500)</td>
<td>51.573</td>
<td>(51.000, 52.300)</td>
</tr>
<tr>
<td>$M_5$</td>
<td>-13.682</td>
<td>(-13.720, -13.580)</td>
<td>52.056</td>
<td>(52.080, 52.120)</td>
</tr>
<tr>
<td>$M_6$</td>
<td>-14.378</td>
<td>(-14.470, -14.370)</td>
<td>51.983</td>
<td>(51.970, 52.020)</td>
</tr>
<tr>
<td>$M_7$</td>
<td>-14.454</td>
<td>(-14.500, -14.430)</td>
<td>51.997</td>
<td>(51.960, 51.980)</td>
</tr>
<tr>
<td>$M_8$</td>
<td>-13.924</td>
<td>(-13.930, -13.830)</td>
<td>52.310</td>
<td>(52.270, 52.320)</td>
</tr>
</tbody>
</table>

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 that is estimated from the option prices. The evidence in favour of option-implied skewness is weaker. However, ignoring 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). The constant volatility models that allow for either excess kurtosis or negative skewness in returns, or both, have the best out-of-sample predictive performance, dominating both the BS and GARCH models. The GARCH models are assigned virtually zero posterior probability when ranked against the constant volatility models. 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.

As a final comment we note 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 data set. In this case, an obvious extension of the methodology outlined in the paper would be to use 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 the predictive associated with any one individual model.

ACKNOWLEDGEMENTS

This research has been supported by an Australian Research Council Large Grant, No. A00103254. The authors would like to thank the Associate Editor and two anonymous referees for providing some very constructive comments on an earlier draft of the paper.

References


27


Notes

1. Lim, Martin and Martin (2004) conduct a classical analysis of option pricing in which the distributional framework of Lye and Martin (1993, 1994) is compared with two alternative approaches, one based on the semi-nonparamtric density of Gallant and Tauchen (1989), and a second based on a mixture of lognormals (see also Melick and Thomas, 1997). The Lye and Martin framework was found to be superior, in general, to the alternative approaches, in terms of fitting and predicting observed option prices.
2. Strictly speaking, the distribution of $C_{ijt}$ should be truncated from below at the no-arbitrage lower bound to be specified in Section 5; see Hull (2000, Chp. 12). However, the incorporation of this truncation in (11) means that $(\beta, \Sigma)$ cannot be integrated out of the joint posterior analytically. In the empirical application we therefore omit the truncation at the estimation stage. However, we do filter the data according to the lower bound, as well as truncate the predictive densities appropriately in the out-of-sample analysis.

3. We have estimated that on a Pentium 4 (Dual Processor, 2Ghz), the time taken to perform the necessary posterior computations would be measured not in hours (or even days), but in weeks, if an MCMC approach to estimation were to be taken.

4. Note that there is a large literature on the market related factors that 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).

**Appendix**

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

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

where $\mu_{t+\Delta t}$ is the conditional mean of the return, $e_{t+\Delta t}$ is an iid 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,

$$
\sigma_{t+\Delta t}^2 = \alpha_0/\Delta t + \alpha_1\sigma_t^2 + \alpha_2\epsilon_t^2,
$$

with $\alpha_0 > 0; \alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 < 1$. Given the discrete time nature of the model in (23) and (24), the Duan (1995, 1999) approach of using an equilibrium model to specify a local risk-neutral valuation measure, is adopted. In the case where $e_{t+\Delta t}$ in (23) is conditionally
normal, the (local) risk-neutral process for returns is defined as
\[
\ln S_{t+\Delta t} - \ln S_t = (\mu_{t+\Delta t} - 0.5 \sigma^2_{t+\Delta t}) \Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t} (z_{t+\Delta t} - \lambda^N_{t+\Delta t}),
\]  
(25)
where \(z_{t+\Delta t}\) is the risk-neutral standard normal innovation and \(\lambda^N_{t+\Delta t}\) is a risk premium given by
\[
\lambda^N_{t+\Delta t} = \sqrt{\Delta t} (\mu_{t+\Delta t} - r_{t+\Delta t})/\sigma_{t+\Delta t},
\]  
(26)
where \(r_{t+\Delta t}\) is the risk-free rate of return. The superscript ‘\(N\)’ in (26) is used to highlight the fact that the risk premium relates specifically to a normal innovation term. Substitution of (26) in (25) produces the following representation of the risk-neutral process,
\[
\ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5 \sigma^2_{t+\Delta t}) \Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t} (z_{t+\Delta t}).
\]  
(27)
The form of the GARCH(1,1) process under local risk-neutralization is then
\[
\sigma^2_{t+\Delta t} = \alpha_0/\Delta t + \alpha_1 \sigma^2_{t+\Delta t} \left( z_{t} - \lambda^N_{t} \right)^2 + \alpha_2 \sigma^2_{t},
\]  
(28)
which produces an unconditional (annualized) variance equal to
\[
\frac{\alpha_0/\Delta t + \alpha_1 \Delta t (\mu_t - r_t)^2}{1 - (\alpha_1 + \alpha_2)}.
\]  
(29)
That is, given \(\alpha_1 > 0\) and \(\mu_t \neq r_t\), local risk-neutralization implies that options are priced under a distribution with a higher unconditional variance than that associated with the objective process in (24).

In order to allow for an innovation term in (23) that accommodates skewness and leptokurtosis, the appropriate risk-neutral model for returns becomes
\[
\ln S_{t+\Delta t} - \ln S_t = (\mu_{t+\Delta t} - 0.5 \sigma^2_{t+\Delta t}) \Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t} \Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}),
\]  
(30)
where \(\Psi^{-1}\) denotes the function that transforms the normal variate, \(z_{t+\Delta t}\), into the relevant nonnormal variate and the risk premium \(\lambda_t\), is now the solution to
\[
E[\Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}) | \mathcal{F}_t] = -\lambda^N_{t+\Delta t},
\]  
(31)
with \(\mathcal{F}_t\) the set of all information up to time \(t\); see Duan (1999) and Hafner and Herwartz (2001). The process for \(\sigma^2_t\) under this so-called generalized local risk-neutral valuation, in turn, becomes
\[
\sigma^2_{t+\Delta t} = \alpha_0/\Delta t + \alpha_1 \sigma^2_t [\Psi^{-1}(z_t - \lambda_t)]^2 + \alpha_2 \sigma^2_t.
\]  
(32)
To implement the risk-neutral adjustments in (30) to (32) 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 (31) as a sample mean; and 4) numerical solution to (31) 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 (30) and (32), 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 inference. Note that 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.

To circumvent these computational problems (30) is rewritten, using (26), as

\[
\ln S_{t+\Delta t} - \ln S_t = (r_{t+\Delta t} - 0.5\sigma^2_{t+\Delta t})\Delta t + \sigma_{t+\Delta t}\sqrt{\Delta t} \left[ \Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}) + \lambda^N_{t+\Delta t} \right]
\]

where

\[
v_{t+\Delta t} = \left[ \Psi^{-1}(z_{t+\Delta t} - \lambda_{t+\Delta t}) + \lambda^N_{t+\Delta t} \right]
\]

is the nonnormal risk-neutral random error term, with conditional mean of zero, given (31). This representation of \( v_{t+\Delta t} \) in (33) and (34) suggests that it can be parameterized directly using a standardized nonnormal distribution, with the parameters of this distribution representing the risk-neutralized parameters. From (32), the risk-neutral process for \( \sigma^2_t \) becomes

\[
\sigma^2_{t+\Delta t} = \frac{\alpha_0}{\Delta t} + \alpha_1 \sigma^2_t \left( v_t - \lambda^N_t \right)^2 + \alpha_2 \sigma^2_t.
\]

The models in (33) and (35) respectively correspond to the models (5) and (6) in the text. As made clear in the text, the particular distributional specification adopted for \( v_t \) nests the normal distribution, in which case \( \lambda_{t+\Delta t} = \lambda^N_{t+\Delta t}, \Psi^{-1} = I, v_{t+\Delta t} = z_{t+\Delta t} \), and the processes in (33) and (35) collapse respectively to those in (27) and (28). In the 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 (33) reduces to

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

With normality, (33) in turn collapses to the discrete version of the risk-neutral distribution that underlies the BS option price, given by (4) in the text.