PRICING AUSTRALIAN S&P200 OPTIONS: A BAYESIAN APPROACH
BASED ON GENERALIZED DISTRIBUTIONAL FORMS\textsuperscript{1}

DAVID B. FLYNN, SIMONE D. GROSE AND GAEL M. MARTIN\textsuperscript{2}

\emph{Monash University}

VANCE L. MARTIN\textsuperscript{3}

\emph{University of Melbourne}

March 18, 2004

\textsuperscript{1}This research has been supported by an Australian Research Council Large Grant and a Monash Small Grant.
\textsuperscript{2}Department of Econometrics and Business Statistics, Monash University, Clayton, 3168, Australia. Corresponding author: Gael Martin; email: gael.martin@buseco.monash.edu.au.
\textsuperscript{3}Department of Economics, University of Melbourne.
Abstract

A new class of option price models is developed and applied to options on the Australian S&P200 Index. The class of models generalizes the traditional Black-Scholes framework by accommodating time-varying conditional volatility, skewness and excess kurtosis in the underlying returns process. An important property of the more general pricing models is that the computational requirements are essentially the same as those associated with the Black-Scholes model, with both methods being based on one-dimensional integrals. Bayesian inferential methods are used to evaluate a range of models nested in the general framework, using observed market option prices. The evaluation is based on posterior parameter distributions, as well as posterior model probabilities. Various fit and predictive measures, plus implied volatility graphs, are also used to rank the alternative models. The empirical results provide evidence that time-varying volatility, leptokurtosis and a small degree of negative skewness are priced in Australian stock market options.

Keywords: Bayesian Option Pricing; Leptokurtosis; Skewness; Time-Varying Volatility; Option Price Prediction; Implied Volatility Smiles
1 Introduction

The Black & Scholes (1973) model for pricing options is founded on two important assumptions: first, that the distribution of returns on the asset on which the option is written is normal, and second, that the volatility of returns is constant. In practice, research on a variety of financial returns has shown that at least one, if not both, of these assumptions are usually violated (for reviews of the relevant literature see Bollerslev, Chou & Kroner, 1992 and Pagan, 1996). This misspecification of the Black-Scholes (BS) model is associated with the occurrence of implied volatility smiles, or skews, in which the volatility implied by equating observed option prices with the BS price, varies with the strike price associated with the option contract; see Hull (2000, Chap. 17). In Bakshi, Chao & Chen (1997), Corrado & Su (1997), Hafner & Herwartz (2001) and Lim, Martin & Martin (2003), amongst others, these implied volatility patterns are directly linked to deviations in the underlying returns process from the BS specifications.

In this paper we develop a more general framework for pricing options that accommodates the empirical features of the underlying returns process, namely time-varying conditional volatility as well as conditional skewness and excess kurtosis. We use a combination of the distributional frameworks of Lye & Martin (1993, 1994) and Fernandez & Steele (1998), augmented with the time-varying volatility specification of Rosenberg & Engle (1997) and Rosenberg (1998). As in Lim et al. (2003), a distribution is specified for the return over the life of the option, with the option then priced by evaluating the expected payoff using simple univariate numerical quadrature. The computational burden is therefore comparable to that associated with the BS price, which requires evaluation of a one-dimensional normal integral. It is also a viable alternative to approaches based on the assumption of stochastic volatility in returns (with or without random jumps), which produce closed form solutions for the option price up to one-dimensional integrals in the complex plane; see, for example Heston (1993), Bates (2000), Chernov & Ghysels (2000) and Pan (2002). Most notably, this approach has distinct computational advantages over the traditional Monte Carlo methods used to price non-BS options, which involve evaluating the expected payoff as an average over many simulation paths; see Hafner & Herwartz (2001), Bauwens & Lubrano (2002) and Martin, Forbes & Martin (2003), amongst others.
The parameters of the alternative models are estimated using observed option prices. That is, ‘implicit’ estimation of the underlying returns models is conducted, contrasting with the alternative approach of estimating the models directly using historical returns data and pricing options via the returns-based parameter estimates. In particular, since observed option prices factor in risk premia, the implicit approach produces estimates of the parameters of the risk-neutral distributions. The inferential approach adopted is Bayesian, with posterior parameter distributions, posterior model probabilities and predictive densities estimated from the option price data. For other recent applications of the Bayesian paradigm to option pricing, see Jacquier & Jarrow (2000), Polson & Stroud (2002), Eraker (2003), Jones (2003) and Martin et al. (2003). The methodology is applied to options written on the Australian S&P200 stock index, the dataset comprising intraday transactions data on all option trades from 14 February 2001 until 31 May 2002. The empirical results thus provide insight into the distributional assumptions that option market participants have factored into their pricing regarding returns on the Australian stock market.

The structure of the paper is as follows. In Section 2 the framework adopted for pricing options is introduced, with the set of alternative models for returns outlined. The simplicity of the evaluation method in the non-BS cases is highlighted. Section 3 describes the Bayesian inferential approach being applied, including the computational details associated with estimation of the marginal posterior densities, model probabilities, predictive densities and implied volatility graphs. The empirical application of the methodology is described in Section 4, with Bayesian model selection criteria supplemented by classical fit and prediction measures. The posterior parameter estimates and model probabilities provide clear evidence that the option market has factored in the assumption of non-constant volatility in returns, plus excess kurtosis in the conditional distribution. Negative skewness in returns is given some support by the data. There is also evidence that the non-BS models fit observed market option prices more accurately than does the BS model, as well as better predicting future prices. The extent of the implied volatility skew is reduced by parameterization of the higher order moments in returns.
2 General Option Pricing Framework

An option is a derivative asset that gives one the right to either buy or sell one unit of the underlying asset at some time in the future, at a prespecified strike or exercise price, $K$. In this paper we focus on the European call option, which gives one the right to buy one unit of the underlying asset at price $K$ when the option matures at time $T$. As such, the value of the European call is a direct function of the price expected to prevail in the spot market for the asset at time $T$. Formally, for a non-dividend paying asset the option price, $q$, is the expected value of the discounted payoff of the option (see Hull, 2000, Chap. 11),

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

where:

$T$ = the time at which the option is to be exercised;

$\tau$ = the length of the option contract, expressed as a proportion of a year;

$K$ = the strike, or 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; and

$E_t (\cdot)$ denotes the conditional expectation, based on information at time $t$, taken with respect to the risk-neutral probability distribution for $S_T$.

From (1), the value of an option at time $t$ is dependent on the known quantities $r$, $K$ and $\tau$, and the observed level of the spot price prevailing at time $t$, $S_t$. The expectation in (1) is evaluated with respect to the risk-neutral probability distribution for $S_T$, this being the pertinent distribution to use in evaluating an option under the assumption of a replicating risk-free portfolio based on the option and the underlying asset; see Hull (2000, Chap. 11). This can be made explicit by rewriting (1) as

$$q = e^{-r \tau} \int_K^{\infty} (S_T - K) g(S_T|S_t) dS_T,$$  \hspace{1cm} (2)

where the function $g(S_T|S_t)$ 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
Accordingly, the continuously compounded return over the life of the option, 
\( \ln(S_T/S_t) \), is assumed to be generated according to

\[
\ln \left( \frac{S_T}{S_t} \right) = \left( r - \frac{1}{2} \sigma_{T|t}^2 \right) \tau + \sigma_{T|t} \sqrt{\tau} e_T, \tag{3}
\]

where \( \sigma_{T|t} \) is the (possibly time-varying) annualized conditional volatility of the return, and \( e_T \) is a random variable such that \( \text{E}(e_T) = 0 \) and \( \text{var}(e_T) = 1 \). The adoption, in (3), of a mean rate of return equal to the risk free rate \( r \) follows from the use of the risk-neutral probability distribution to evaluate the option.

To relax the constant volatility assumption we adopt the conditional volatility specification of Rosenberg & Engle (1997) and Rosenberg (1998), in which volatility is specified as a function of the net return over the life of the option,

\[
\sigma_{T|t} = \exp \left( \delta_1 + \delta_2 \ln \left( \frac{S_T}{S_t} \right) \right). \tag{4}
\]

This particular representation for \( \sigma_{T|t} \) renders the latter both a time-varying function, via the dependence on \( S_t \), and a random function, via the dependence on \( S_T \). The conditional pdf of \( S_T \) then follows as

\[
g(S_T|S_t) = |J| p(e_T), \tag{5}
\]

where \( p(e_T) \) is the pdf of \( e_T \), \(|J| \) is the Jacobian of the transformation from \( e_T \) to \( S_T \), and

\[
J = \frac{\partial e_T}{\partial S_T} = \frac{1}{S_T \sigma_{T|t} \sqrt{\tau}} \left( 1 + \delta_2 \sigma_{T|t}^2 \tau - \delta_2 \left( \ln \left( \frac{S_T}{S_t} \right) - \left( r - \frac{1}{2} \sigma_{T|t}^2 \right) \tau \right) \right).
\]

Several alternative distributional assumptions are adopted for \( e_T \), implying, via (5), several alternative assumptions about the form of \( g(S_T|S_t) \) and, hence, about the value of the theoretical option price in (2). Following Fernandez & Steel (1998), we begin by defining a random variable \( W \) with finite variance and pdf, \( f(\cdot) \), symmetric around a single mode at zero. We then denote by \( W_\gamma \), the random variable with pdf,

\[
f_\gamma(w) = \frac{2}{\gamma + \frac{1}{\gamma}} \left( f \left( \frac{w}{\gamma} \right) I_{[0,\infty)}(w) + f \left( \gamma w \right) I_{(-\infty,0)}(w) \right), \tag{6}
\]

where \( I_A(\cdot) \) denotes the indicator function for the set \( A \), and the mean and variance of \( W_\gamma \) are respectively

\[
\text{E}(W_\gamma) = \mu_\gamma = (\gamma - \frac{1}{\gamma}) \text{E}(|W|)
\]
and
\[ \text{var}(W_\gamma) = \sigma_\gamma^2 = (\gamma^2 + \frac{1}{\gamma^2} - 1)E(W^2) - \mu_\gamma^2. \]

The parameter \( \gamma \) introduces skewness into the distribution of \( W_\gamma \), with \( \gamma > 1 \) producing positive skewness and \( \gamma < 1 \) negative skewness. When \( \gamma = 1 \), the pdf defined in (6) is symmetric, with \( f_\gamma(\cdot) = f(\cdot) \) and \( W_\gamma \equiv W \). Moreover, if \( f(\cdot) \) in (6) is defined as a pdf with excess kurtosis, (6) produces a distribution for \( W_\gamma \) with both leptokurtosis and skewness, with the symmetric leptokurtic distribution recovered by setting \( \gamma = 1 \).

In specifying a leptokurtic form for \( f(\cdot) \) we use a subordinate distribution from the generalized exponential family defined in Lye & Martin (1993, 1994) that has excess kurtosis relative to the normal distribution, namely
\[ f(w) = \kappa \left( 1 + \frac{w^2}{\nu} \right)^{-\frac{1}{2}(\nu+1)} \exp \left( -\frac{1}{2}w^2 \right), \tag{7} \]

where \( \kappa \) must be computed numerically. The standardized variable \( e_T \) in (3) is then defined by
\[ e_T \overset{d}{=} \frac{W_\gamma - \mu_\gamma}{\sigma_\gamma}, \tag{8} \]

with pdf
\[ p(e_T) = \sigma_\gamma f_\gamma \left( \mu_\gamma + \sigma_\gamma e_T \right). \tag{9} \]

The pdf in (7) is proportional to a product of Student \( t \) and normal kernels, with the first term allowing for kurtosis in excess of that of the normal distribution for any finite value of \( \nu \), and the second term ensuring that the moments of \( W_\gamma \), and hence \( e_T \), exist for any value of \( \nu > 0 \). The second term also ensures that the moments of \( S_T \) take with respect to the density in (5) exist for all \( \nu > 0 \), so that that the option price in (2) is well-defined. This is in contrast with the lack of moment existence for \( S_T \) that would be associated with the specification of a Student \( t \) distribution for the continuously compounded return; see Duan (1999). The degree of kurtosis in \( p(e_T) \) has no closed form solution and so an estimate of the kurtosis coefficient, \( E(e_T^4) \), must be computed numerically. Similarly, the Pearson skewness coefficient associated with the standardized variate \( e_T \), \( E(e_T^3) \), can be estimated numerically. We refer to the pdf in (7) as a generalized Student \( t \) (GST) density.

In summary, the general framework proposed here nests 8 alternative models for the standardized variate \( e_T \), denoted respectively by \( M_1 \) to \( M_8 \). Each model, including
its associated parameter set, is listed in Table 1. The models for which \( \delta_2 \) in (4) is set to zero (the first panel of models in Table 1) are models for which constant volatility is imposed, but with skewness and/or leptokurtosis in returns accommodated. The models for which \( \delta_2 \) is estimated freely (listed in the second panel in Table 1) allow for both time-varying volatility and conditional skewness and leptokurtosis. The distribution of \( e_T \) is defined by (6) to (9), with the pdf \( f(\cdot) \) in (6) specified respectively as normal and GST. When \( \gamma = 1 \), the distribution of \( e_T \) is symmetric; otherwise it is skewed. Model \( M_1 \) corresponds to the BS specifications, in which case \( g(S_T|S_t) \) is log-normal and the integral in (2) has the well-known solution,

\[
q = S_t\Phi(d_1) - K^{-rT}\Phi(d_2),
\]

where

\[
d_1 = \frac{1}{\sigma\sqrt{T}}\left(\ln\frac{S_t}{K} + (r + \frac{1}{2}\sigma^2)T\right),
\]

\[
d_2 = \frac{1}{\sigma\sqrt{T}}\left(\ln\frac{S_t}{K} + (r - \frac{1}{2}\sigma^2)T\right),
\]

\( \sigma_{T|h} = \sigma = \exp(\delta_1) \) is the constant volatility parameter, and \( \Phi(\cdot) \) denotes the standard normal cumulative distribution function. For all other specifications, the theoretical price in (2) is calculated by evaluating the relevant integral using one-dimensional numerical quadrature.

### Table 1:
Parameterization of Alternative Models Based on Equations (3) to (9)

<table>
<thead>
<tr>
<th>( p(e_T) )</th>
<th>Constant volatility</th>
<th>Time-varying volatility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>( M_1 (\delta_1, \delta_2 = 0, \gamma = 1) )</td>
<td>( M_5 (\delta_1, \delta_2, \gamma = 1) )</td>
</tr>
<tr>
<td>Skewed Normal</td>
<td>( M_2 (\delta_1, \delta_2 = 0, \gamma) )</td>
<td>( M_6 (\delta_1, \delta_2, \gamma) )</td>
</tr>
<tr>
<td>GST</td>
<td>( M_3 (\delta_1, \delta_2 = 0, \gamma = 1, \nu) )</td>
<td>( M_7 (\delta_1, \delta_2, \gamma = 1, \nu) )</td>
</tr>
<tr>
<td>Skewed GST</td>
<td>( M_4 (\delta_1, \delta_2 = 0, \gamma, \nu) )</td>
<td>( M_8 (\delta_1, \delta_2, \gamma, \nu) )</td>
</tr>
</tbody>
</table>
3 Bayesian Inference in an Option Pricing Framework

3.1 Posterior Density Functions

Assuming the theoretical option pricing function (2) is an unbiased representation of the true data generation process, we can specify a model for the \( i \)th observed option price, \( C_i \), as

\[ C_i = q_i(z_i, \theta_k) + u_i, \quad i = 1, \ldots, N, \]  

where \( q_i(\cdot, \cdot) \) denotes the \( i \)th theoretical price function as per (2), \( z_i = (r_i, K_i, \tau_i, S_i) \) represents the known set of factors needed to calculate \( q_i(\cdot, \cdot) \), \( u_i \) is the random pricing error, and \( \theta_k \) is the vector of unknown parameters that characterize the \( k \)th model for the underlying returns distribution, \( M_k \), \( k = 1, 2, \ldots, 8 \). The index \( i \) represents variation over time as well as variation across different option contracts at any given point in time, with \( N \) denoting the number of option prices in the sample.

The inclusion of a stochastic error term in (11) serves as recognition of the fact that option pricing models are only approximations of the true underlying process driving observed prices. The inevitable pricing error derives at least in part from ‘model’ error; that is, the model being used to calculate the theoretical price is incorrect either in its specification or in the values assumed for its parameters. The error may also arise via the non-synchronous recording of spot and option prices, transaction costs and other market frictions. The possibility of a systemic component in the pricing error can be accommodated by extending (11) to include a regression component. In its simplest form this implies

\[ C_i = \beta_1 + \beta_2 q_i(z_i, \theta_k) + u_i, \quad i = 1, \ldots, N. \]  

The addition of the following distributional assumption for \( u_i \),

\[ u_i \overset{d}{=} N(0, \sigma_u^2) \quad \text{for all} \quad i = 1, \ldots, N, \]  

with \( u_i \) and \( u_j \) independent for all \( i \neq j \), then yields the model adopted in Section 4. As demonstrated in Bates (2000), Eraker (2003) and Martin et al. (2003), more general specifications are possible for both (12) and (13). In addition, \( u_i \) should, strictly speaking, be truncated from below according to the ‘no-arbitrage’ lower bound.
for $C_i$,

$$C_i \geq \lambda_i = \max\{0, S_i - e^{-r_i T_i} K_i\}, \quad (14)$$

(see Hull, 2000, Sect. 7.3). However, such truncation adds considerably to the computational burden, whilst having only a negligible impact on the parameter estimates. Instead the data is simply filtered according to the lower bound, as well as the lower bound being imposed when producing predictive densities for out-of-sample prices.

The distributional assumptions in (12) and (13) define the likelihood function,

$$L(\theta_k, \beta, \sigma_u|c) = (2\pi)^{-\frac{N}{2}}\sigma_u^N \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{C_i - x_i(\theta_k)^\top \beta}{\sigma_u} \right)^2 \right\}, \quad (15)$$

where $\beta = (\beta_1, \beta_2)$, $x_i(\theta_k) = (1, q_i(z_i, \theta_k))$ and $c = (C_1, C_2, \ldots, C_N)$ is the $(N \times 1)$ vector of observed option prices. In order to simplify the notation, we emphasize the dependence of $x_i(\theta_k)$ on the unknown parameter vector $\theta_k$, whilst not making explicit its dependence on the known factors in $z_i$. We likewise choose not to make explicit the dependence of the likelihood function on $z_i$, $i = 1, 2, \ldots, N$.

Defining the full set of unknown parameters for model $M_k$ as $\omega_k = \{\beta, \sigma_u, \theta_k\}$, and assuming a priori independence between $\beta, \sigma_u$ and $\theta_k$ respectively, we define a prior for $\omega_k$ of the form

$$p(\omega_k) \propto \frac{1}{\sigma_u} p(\theta_k). \quad (16)$$

Details of $p(\theta_k)$ are provided in Section 4.2. The prior on $\{\beta, \sigma_u\}$ is the standard noninformative prior for the location and scale parameters in a regression model.

Given (15) and (16), the joint posterior pdf for $\omega_k$ is given by

$$p(\omega_k|c) \propto \sigma_u^{-(N+1)} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{C_i - x_i(\theta_k)^\top \beta}{\sigma_u} \right)^2 \right\} p(\theta_k). \quad (17)$$

The parameters $\{\beta, \sigma_u\}$ can be integrated out of $p(\omega_k|c)$ using standard analytical results, resulting in a marginal posterior for $\theta_k$ of the form,

$$p(\theta_k|c) \propto \hat{\sigma}_u^{-(N-2)} |X(\theta_k)^\top X(\theta_k)|^{-\frac{1}{2}} p(\theta_k), \quad (18)$$

where

$$\hat{\sigma}_u^2 = \frac{(C_i - x_i(\theta_k)^\top \hat{\beta})^\top (C_i - x_i(\theta_k)^\top \hat{\beta})}{N-2},$$

$$\hat{\beta} = (X(\theta_k)^\top X(\theta_k))^{-1} X(\theta_k)^\top c,$$
and $X(\theta_k)$ is the $(N \times 2)$ matrix $X(\theta_k) = [1 \ \ q(\theta_k)]$, with 1 the $N$-vector of ones, and $q(\theta_k) = (q_1(z_1, \theta_k), q_2(z_2, \theta_k), \ldots, q_N(z_N, \theta_k))$. As $\theta_k$ is of low dimension for all models considered, the marginal posterior densities for the individual elements of $\theta_k$ are obtained by applying deterministic numerical integration methods to the density in (18). The marginal posterior pdf’s for $\beta$ and $\sigma_u$ then follow as

$$p(\beta | c) = \int p(\beta | \theta_k, c)p(\theta_k | c)d\theta_k$$

$$\propto \int_{\theta_k} \left(1 + \frac{(\beta - \hat{\beta})X(\theta_k)^T X(\theta_k)(\beta - \hat{\beta})}{(N - 2)\hat{\sigma}_u^2}\right)^{-\frac{N}{2}} p(\theta_k | c)d\theta_k \tag{19}$$

and

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

$$\propto \int_{\theta_k} \sigma_u^{-(N-1)} \exp \left(-\frac{(N - 2)\hat{\sigma}_u^2}{2\sigma_u^2}\right) p(\theta_k | c)d\theta_k. \tag{20}$$

### 3.2 Posterior Model Probabilities

To determine the model that is most probable given the information in the observed option prices, we construct implicit model probabilities for each of the models $M_1, M_2, \ldots, M_8$ and rank them in order of highest to lowest value. The model probabilities are constructed via the estimation of posterior odds ratios for the models $M_2, M_3, \ldots, M_8$, relative to reference model $M_1$. The posterior odds ratio for model $M_k$ relative to $M_1$, $O_{k1}$, is given by the product of the prior odds ratio, $\Pr(M_k)/\Pr(M_1)$, and the Bayes factor, $p(c|M_k)/p(c|M_1)$,

$$O_{k1} = \frac{\Pr(M_k|c)}{\Pr(M_1|c)} = \frac{\Pr(M_k)}{\Pr(M_1)} \times \frac{p(c|M_k)}{p(c|M_1)}, \quad k = 2, 3, \ldots, 8, \tag{21}$$

where $\Pr(M_k|c)$ is the posterior probability of model $M_k$, and

$$p(c|M_k) = \int_{\omega_k} p(c|\omega_k, M_k)p(\omega_k|M_k)d\omega_k$$

$$= \int_{\omega_k} L(\omega_k|M_k)p(\omega_k|M_k)d\omega_k \tag{22}$$

is the marginal likelihood of model $M_k$. Since $\beta$ and $\sigma_u$ can be integrated out of (17) analytically, it follows that

$$p(c|M_k) = h \int_{\theta_k} L^*(\theta_k|M_k)p(\theta_k|M_k)d\theta_k, \tag{23}$$
where \( L^*(\theta_k|M_k) = \sigma_u^{-(N-2)} \left| X(\theta_k)^\top X(\theta_k) \right|^{-1/2} \) and \( h \) is a constant that is not dependent on the specification of \( M_k \). Conveniently, because the integral in (23) is just the integrating constant required for the normalization of (18) above, computation of the marginal likelihood for each model requires nothing over and above that required to produce the marginal posterior densities for each \( \theta_k \). (cf. Chib, 1995). The posterior model probability for model \( M_k \) then follows as

\[
Pr(M_k|c) = \frac{p(c|M_k) Pr(M_k)}{\sum_{j=1}^{S} p(c|M_j) Pr(M_j)} = \frac{O_{k1}}{\sum_{j=1}^{S} O_{j1}}, \quad k = 2, 3, ..., 8.
\]

### 3.3 Predictive Density Functions

Whilst the posterior model probabilities are a Bayesian measure of the performance of the alternative models within-sample, the relative out-of-sample performance of the models can be assessed via the ability of each model to predict future option prices accurately. Having specified a returns model \( M_k \) indexed on \( \theta_k \), the predictive pdf for an option price \( C_f \) associated with some future time period \( f \), given model \( M_k \) and the data, is

\[
p(C_f|c) = \int_{\sigma_u} \int_{\beta} \int_{\theta_k} p(C_f, \beta, \sigma_u, \theta_k|c) d\theta_k d\beta d\sigma_u
\]

\[
= \int_{\theta_k} p(C_f|\theta_k, c) p(\theta_k|c) d\theta_k,
\]

where

\[
p(C_f|\theta_k, c) \stackrel{d}{=} \mu_{fk} + \sigma_{fk} t_{\nu},
\]

\[
\mu_{fk} = x_f(\theta_k)^\top \hat{\beta},
\]

\[
\sigma_{fk}^2 = \sigma_u^2 \left(1 + x_f(\theta_k)^\top (X(\theta_k)^\top X(\theta_k))^{-1} x_f(\theta_k)\right),
\]

\( t_{\nu} \) denotes the Student \( t \) variate with \( \nu \) degrees of freedom, and \( x_f(\theta_k) = (1, q_f(\mathbf{z}_{t_f}, \theta_k)) \). The vector \( \mathbf{z}_f = (r_f, K_f, \tau_f, S_f) \) encompasses the known set of factors needed to calculate \( q_f(\cdot, \cdot) \).

The predictive pdf in (25) can therefore be estimated by computing a weighted average of the conditional densities for \( C_f \), \( p(C_f|\theta_k, c) \), with the weights being the probability mass assigned to each gridpoint in the numerically evaluated posterior
density, \( p(\theta_k|c) \, d\theta_k \). That is, the predictive pdf is estimated as

\[
\hat{p}(C_f|c) = w \sum_{j=1}^{N_{\theta_k}} p(C_f|\theta_k^{(j)}, c)p(\theta_k^{(j)}|c),
\]

(26)

where \( N_{\theta_k} \) is the number of gridpoints used in defining the density \( p(\theta_k|c) \), \( w \) is the grid width and \( p(\theta_k^{(j)}|c) \) denotes the ordinate of \( p(\theta_k|c) \) at grid value \( \theta_k^{(j)} \). The density \( p(C_f|\theta_k^{(j)}, c) \) is truncated at the no-arbitrage lower bound (14) and renormalized prior to computing the weighted sum in (26).

### 3.4 Implied Volatility Graphs

The \( i \)th option contract is said to be *in-the-money (ITM)* if its immediate exercise would lead to a positive cash flow, that is, if the current value of the spot price, \( S_i \), exceeds the value of the strike price, \( K_i \). Similarly, the option is *out-of-the-money (OTM)* if \( S_i \) is less than \( K_i \), and *at-the-money (ATM)* if \( S_i \) and \( K_i \) are equal.

Measuring the degree of moneyness by the ratio, \( S_i/K_i \), variation in implied volatility across different values for this ratio (or across different values for the strike price \( K_i \) for a fixed spot price \( S_i \)) is generally taken to indicate some degree of misspecification of the option pricing model, given that volatility is a feature of the underlying asset returns and not a function of the degree of moneyness of an option written on that asset. In particular, this variation can be directly linked to the fact that participants in the options market have factored into their pricing, distributional assumptions about the underlying returns process that deviate from the BS specifications; see, for example, Hafner & Herwartz (2001) and Lim et al. (2003).

Accordingly, yet another criterion for testing a model’s ability to characterize option prices is the relative flatness of its implied volatility graph. That is, for different values for \( S_i/K_i \), the volatility implied by each model is calculated by setting the \( i \)th theoretical price equal to the corresponding observed price, then solving for the volatility parameter. In the case of the BS model this is straightforward, as the volatility is the only unknown parameter that characterizes the theoretical BS price. All other models have additional distributional parameters whose values are most conveniently set equal to their estimated marginal posterior modes. The model deemed to best characterize option price behaviour is the one that generates the flattest implied volatility graph. As the implied volatility graphs associated with the BS
model typically depict certain stylized shapes, such as ‘smiles’, ‘smirks’, ‘frowns’ or ‘skews’, the production of a flatter graph via the use of a non-BS model, is sometimes referred to as a reduction, or correction of the volatility skew.

4 Pricing Options on the S&P200

4.1 Data

The methodology outlined above is applied to the intraday transaction prices of call options written on the Australian S&P200 Index, observed over the period 14 February 2001 to 31 May 2002. The S&P200 Index represents the ‘price’ of a market portfolio covering 200 of the largest companies trading on the Australian Stock Exchange (ASX), comprising approximately 89% of the total market capitalization of the Australian stock market (as at 31 August 2000). Options on the Australian S&P200 are European style options, expiring at 3-monthly intervals. Settlement at exercise is in cash.

Each transaction record includes the contract date and time, the option price \( C_i \), the strike price \( K_i \) and the time to maturity \( \tau_i \), but not the value of the underlying index at the time the contract was written. The latter information is extracted from data on the S&P200 index itself, recorded at one minute intervals up to 18 January 2002, and at approximately 30 second intervals thereafter. The relevant ‘spot’ price, \( S_i \), is then taken to be the recorded value of the Index most closely synchronized with the \( i \)th transaction, discounted by the average dividend rate; see Hull (2000, Sect. 11.12). That is, the Index value associated with each observational point, \( S_i \), is replaced by the discounted value, \( S_i e^{-d\tau_i} \), where \( d = 0.033 \) is the annualized average dividend rate based on total dividends paid over the sample period.

After excluding option contracts with very long times to maturity, several observations that appear to be in error, and those prices that do not satisfy the arbitrage restriction (14), the final dataset contains 5471 trades. Time to maturity, \( \tau_i \), ranges from 1/365 to 281/365 across the sample. The moneyness \( (S_i/K_i) \) of the contracts in the sample ranges from 0.81 to 1.18, with the exception of a handful of trades with moneyness in excess of 1.18. The risk-free rate, \( r_i \), prevailing at the time of the contract is deemed to be that day’s interest charged on 90-day Bank Accepted Bills. All option price, Index and dividend data were obtained from the ASX; the data on
interest rates was obtained from the Reserve Bank of Australia Bulletin. We reserve the data for the last 8 days, that is, the trades occurring between 22 May and 31 May 2002 inclusive, to assess the out-of-sample performance of the models. This division of the final dataset into estimation and validation subsamples leaves 5356 in-sample and 115 out-of-sample observations.

Table 2:
Summary of S&P200 Option Price Data

<table>
<thead>
<tr>
<th>Moneyness ($S_i/K_i$)</th>
<th>Number of Trades</th>
<th>Average Market Price (cents/share)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimation period</td>
<td>Validation period</td>
</tr>
<tr>
<td>OTM ($&lt; 0.97$)</td>
<td>2002</td>
<td>29</td>
</tr>
<tr>
<td>ATM (0.97 – 1.03)</td>
<td>2943</td>
<td>86</td>
</tr>
<tr>
<td>ITM ($&gt; 1.03$)</td>
<td>411</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>5356</td>
<td>115</td>
</tr>
</tbody>
</table>

Table 2 provides a summary of the main characteristics of the option price data, with observations divided into three categories according to their level of moneyness, as defined by Bakshi et al. (1997). Specifically, option $i$ is here categorised as ATM if $S_i/K_i \in (0.97, 1.03)$, OTM if $S_i/K_i < 0.97$, and ITM if $S_i/K_i > 1.03$. We note that the number of ITM trades comprises less than 8% of the total overall; with there being none at all in our validation period. In fact there are no in the money options traded after 9 May 2002 that are not excluded by filtering according to (14).

4.2 Priors

As noted in (16) above, the prior pdf on the full set of unknowns for model $M_k$ is defined as proportional to the product of the standard noninformative prior for $\{\beta, \sigma_u\}$ and the prior on the remaining model parameters, $p(\theta_k)$. We assume a priori
independence between all elements of $\theta_k$, as well as specifying priors for all elements that are either noninformative, or informative but diffuse. Such a prior specification ensures that the information in the option price data dominates the posterior results.

More specifically, uniform priors are adopted for $\delta_1$ and $\delta_2$. The uniform prior for $\delta_1$ implies a prior for $\sigma = \exp(\delta_1)$ in the constant volatility ($\delta_2 = 0$) models proportional to $1/\sigma$. The prior pdf for the degrees of freedom parameter (if present) is taken to be exponential, $p(\nu) = \lambda e^{-\lambda \nu}$, with a hazard rate $\lambda = 0.1$. This implies a prior mean for $\nu$ equal to 10 and prior variance equal to 100. This essentially produces a prior distribution that is half-way between one having normal tails and one having fat tails (see Fernandez & Steel, 1998). For the skewness parameter we again follow Fernandez & Steel by specifying a gamma prior on $\gamma^2$ with values of the scale and shape parameters that imply that $E(\gamma) = 1$, $\text{var}(\gamma) = 0.57$, and $\Pr(\gamma < 1) = 0.58$. Both of these latter priors, whilst proper, are quite diffuse, with preliminary analysis indicating that the posterior parameter estimates reported in Tables 3 and 4 differ very little from those produced using uniform priors for both $\nu$ and $\gamma$. Using the criterion suggested in Kass (1993, p.557), the model probabilities reported in Table 5 have also been found to be robust to the prior specifications for $\nu$ and $\gamma$. Details of these robustness results can be obtained from the authors on request.

Since $\delta_1$ is a parameter common to all models, the specification of an improper prior for this parameter does not introduce any arbitrariness in the construction of the posterior odds ratio. That is, the undefined integrating constant for the marginal prior of $\delta_1$ effectively cancels in the ratio of marginal likelihoods in (21). On the other hand, the uniform prior for $\delta_2$ is, strictly speaking, invalid for use in the posterior odds ratio in (21) when $M_k$ corresponds to any of the models for which $\delta_2 \neq 0$, since the marginal likelihood calculation in (23) has an arbitrariness associated with it, depending on the range over which $\delta_2$ is integrated. Nevertheless, in practice, with a sample size as large as that used in the empirical application, it is straightforward to determine a range of integration over which the likelihood function has virtually all mass, thereby eliminating the arbitrary aspect of the calculation in (23). Moreover, assessment of the sensitivity of the model probabilities to the integration range for $\delta_2$, again using the criterion of Kass (1993), indicates that the impact is negligible.
4.3 Empirical Results

4.3.1 Posterior Parameter Estimates

The joint posterior pdf for the parameters of each of the option pricing models is estimated by setting up a grid of parameter values, and evaluating the kernel of the posterior at each gridpoint. In order to avoid numerical overflow problems, we compute the log-kernel at each of the gridpoints, from which the largest ordinate value is subtracted. When exponentiated, these scaled posterior ordinates are then rescaled using numerical integration so as to produce a normalized joint density function, as described in Section 3.1. Each marginal posterior is produced by further applications of numerical integration.

As highlighted earlier, the proposed method for evaluating any of the non-BS option prices is competitive with evaluation of the BS price, requiring as it does the simple application of univariate numerical integration. As such, for given parameter values, the computational requirements of the more general pricing method are essentially identical to those required for BS pricing. At the level of estimation, however, the non-BS models require more computation time than the BS model simply as a result of their being more highly parameterized. That said, the non-BS models proposed here are quite parsimonious, with the unknown parameters not exceeding four in number. As a consequence, the use of deterministic integration methods for evaluating the posterior density functions is feasible, with computing times, in the case of the sample used in the empirical application, ranging from minutes to a number of hours on a 2 GHz Pentium 4, depending on the precise specification of the model concerned, and the fineness of the grid used to estimate the posterior pdf. More highly parameterized models would be more efficiently analysed using a Markov chain Monte Carlo simulation algorithm.

Summary measures of the marginal posterior pdf’s for the constant volatility ($\delta_2 = 0$) and time-varying volatility ($\delta_2 \neq 0$) models respectively are reported in Tables 3 and 4. The measures comprise marginal posterior means and modes, plus 95% Highest Posterior Density (HPD) intervals. Considering the results in Table 4, it is clear that the option prices have factored in the assumption of time-varying volatility, with both point and interval estimates of $\delta_2$ indicating a non-zero value for this parameter. The point estimates of $\delta_1$ indicate an estimate of volatility at the time of maturity (at
Table 3:
Marginal Posterior Means and Modes, plus 95% HPD Intervals:
Constant Volatility Models: $\delta_2 = 0; \sigma = \exp(\delta_1)$

<table>
<thead>
<tr>
<th>Model</th>
<th>$p(e_T)$</th>
<th>Parameter</th>
<th>Marginal Mode</th>
<th>Marginal Mean</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>Normal</td>
<td>$\sigma$</td>
<td>0.1192</td>
<td>0.1192</td>
<td>(0.1171, 0.1213)</td>
</tr>
<tr>
<td>$M_2$</td>
<td>Skewed Normal</td>
<td>$\sigma$</td>
<td>0.1192</td>
<td>0.1192</td>
<td>(0.1170, 0.1214)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>0.9980</td>
<td>1.0001</td>
<td>(0.9474, 1.0544)</td>
</tr>
<tr>
<td>$M_3$</td>
<td>GST</td>
<td>$\sigma$</td>
<td>0.1242</td>
<td>0.1242</td>
<td>(0.1214, 0.1270)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>0.7581</td>
<td>0.9480</td>
<td>(0.4292, 1.6428)</td>
</tr>
<tr>
<td>$M_4$</td>
<td>Skewed GST</td>
<td>$\sigma$</td>
<td>0.1252</td>
<td>0.1253</td>
<td>(0.1220, 0.1287)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>0.9656</td>
<td>0.9669</td>
<td>(0.9225, 1.0125)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>0.6606</td>
<td>0.8094</td>
<td>(0.3691, 1.3926)</td>
</tr>
</tbody>
</table>

Table 4:
Marginal Posterior Means and Modes, plus 95% HPD Intervals:
Time-Varying Volatility Models

<table>
<thead>
<tr>
<th>Model</th>
<th>$p(e_T)$</th>
<th>Parameter</th>
<th>Marginal Mode</th>
<th>Marginal Mean</th>
<th>95% HPD Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_5$</td>
<td>Normal</td>
<td>$\delta_1$</td>
<td>-2.1717</td>
<td>-2.1718</td>
<td>(-2.1898, -2.1539)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta_2$</td>
<td>0.3826</td>
<td>0.3681</td>
<td>(0.3267, 0.3975)</td>
</tr>
<tr>
<td>$M_6$</td>
<td>Skewed Normal</td>
<td>$\delta_1$</td>
<td>-2.1682</td>
<td>-2.1684</td>
<td>(-2.1871, -2.1499)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta_2$</td>
<td>0.3901</td>
<td>0.3786</td>
<td>(0.3342, 0.4116)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>0.9550</td>
<td>0.9551</td>
<td>(0.9052, 1.0061)</td>
</tr>
<tr>
<td>$M_7$</td>
<td>GST</td>
<td>$\delta_1$</td>
<td>-2.1376</td>
<td>-2.1374</td>
<td>(-2.1613, -2.1135)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta_2$</td>
<td>0.3367</td>
<td>0.3187</td>
<td>(0.2608, 0.3589)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>1.1256</td>
<td>1.6032</td>
<td>(0.5892, 3.1652)</td>
</tr>
<tr>
<td>$M_8$</td>
<td>Skewed GST</td>
<td>$\delta_1$</td>
<td>-2.1228</td>
<td>-2.1221</td>
<td>(-2.1487, -2.0953)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\delta_2$</td>
<td>0.3453</td>
<td>0.3306</td>
<td>(0.2810, 0.3666)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>0.9359</td>
<td>0.9379</td>
<td>(0.8936, 0.9832)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\nu$</td>
<td>0.8493</td>
<td>1.1380</td>
<td>(0.4493, 2.1173)</td>
</tr>
</tbody>
</table>
which point $S_t = S_T$) of approximately 12% in annualized terms, which tallies closely with the point estimates of $\sigma = \exp(\delta_1)$ in the constant volatility models reported in Table 3. All but one of the point estimates of $\gamma$ in Tables 3 and 4 indicate that a small amount of negative skewness in returns has been factored into the option prices. The negative skewness is further confirmed by the interval estimate of $\gamma$ for $M_8$, which covers values less than unity. However the interval estimates of $\gamma$ for $M_2$, $M_4$ and $M_6$ suggest some uncertainty as to the existence of skewness, covering as they do values for $\gamma$ that are associated with symmetry ($\gamma = 1$), negative skewness ($\gamma < 1$) and positive skewness ($\gamma > 1$).

Most notably, the results in both tables provide clear evidence of option-implied excess kurtosis, both for the conditional returns distributions associated with the models in Table 4 and for the unconditional distributions associated with the constant volatility models in Table 3. For the GST models the point estimates of $\nu$ range from 0.66 ($M_4$) to 1.60 ($M_7$), implying a range of kurtosis estimates of 3.75 to 3.49. The interval estimates of $\nu$ in all of the GST models indicate values of kurtosis that represent a departure from the kurtosis of 3 associated with the normal distribution. For the time-varying volatility models that allow for excess kurtosis, namely $M_7$ and $M_8$, the kurtosis estimated is for the conditional distribution of returns. Hence, we would anticipate the smaller degree of kurtosis that is estimated for these models in comparison with the kurtosis estimated for the corresponding constant volatility models, $M_3$ and $M_4$, since the time-varying volatility specification itself is expected to capture some of the kurtosis in the data.

Finally, the estimates of the constant volatility parameter $\sigma$ range from a low of 11.92% for the BS model ($M_1$) to a high of 12.53% for the skewed GST model ($M_4$). The interval estimates of $\sigma$ for all models are quite narrow, being less that one percentage point in width.

4.3.2 Within Sample Performance

Table 5 reports various criteria for measuring the within-sample performance of the alternative models. In addition to the posterior model probabilities, computed according to (24) with equal prior probabilities for all models, we include the Root
Mean Squared Error (RMSE) measure of fit for the $k$th model,

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( C_i - \left( \hat{\beta}_1 + \hat{\beta}_2 q_i \left( z_i, \hat{\theta}_k \right) \right) \right)^2},$$

where $\hat{\theta}_k$ is the marginal posterior mode for $\theta_k$, and $\hat{\beta}_1$ and $\hat{\beta}_2$ are Ordinary Least Squares estimates of $\beta_1$ and $\beta_2$ respectively, based on $q_i(z_i, \hat{\theta}_k)$. As a fit measure that adjusts for the varying magnitudes of the different option prices in the sample, we also report the Mean Absolute Percentage Error (MAPE),

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{C_i - \left( \hat{\beta}_1 + \hat{\beta}_2 q_i \left( z_i, \hat{\theta}_k \right) \right)}{C_i} \right| \times 100\%.$$

To cater for the different number of unknown parameters across models we report the Akaike Information Criterion (AIC) and Schwarz Information Criterion (SIC) measures of fit, computed as

$$AIC = -\frac{2}{N} \left( \ln \hat{L} - K_k \right), \quad SIC = -\frac{2}{N} \left( \ln \hat{L} - \frac{1}{2} K_k \ln N \right)$$

respectively, where $\hat{L}$ denotes the likelihood function in (15) evaluated at $\hat{\theta}_k, \hat{\beta}_1, \hat{\beta}_2$ and $K_k$ is the number of unknown parameters in model $M_k$, $k = 1, 2, \ldots, 8$.

Table 5:

<table>
<thead>
<tr>
<th>Measures of Within-Sample Performance</th>
</tr>
</thead>
</table>

**Constant Volatility Models**

| Model | RMSE | MAPE | AIC  | SIC   | $Pr(M_k|c)$ |
|-------|------|------|------|-------|------------|
| $M_1$ | 0.1975 | 49.29 | -0.4052 | -0.4015 | 0.0000 |
| $M_2$ | 0.1975 | 49.27 | -0.4048 | -0.3999 | 0.0000 |
| $M_3$ | 0.1970 | 50.10 | -0.4101 | -0.4052 | 0.0000 |
| $M_4$ | 0.1969 | 50.01 | -0.4101 | -0.4039 | 0.0000 |

**Time-Varying Volatility Models**

| Model | RMSE | MAPE | AIC  | SIC   | $Pr(M_k|c)$ |
|-------|------|------|------|-------|------------|
| $M_5$ | 0.1968 | 49.98 | -0.4122 | -0.4073 | 0.0017 |
| $M_6$ | 0.1967 | 48.72 | -0.4124 | -0.4063 | 0.0002 |
| $M_7$ | 0.1965 | 49.66 | -0.4148 | -0.4086 | 0.5670 |
| $M_8$ | 0.1963 | 49.19 | -0.4158 | -0.4085 | 0.4311 |

18
The results in Table 5 substantiate the estimation results presented in Tables 3 and 4. As is evident from a comparison of the results in the two panels in Table 5, the models that allow for the time-varying volatility specification in (4) dominate the corresponding constant volatility models. The models that augment the time-varying volatility specification with a GST conditional distribution ($M_7$ and $M_8$) are together assigned virtually all posterior probability weight in the entire set of 8 models, with the symmetric model ($M_7$) slightly out-performing the skewed model ($M_8$). Between them, these two models also have the lowest RMSE, AIC and SIC values out of all 8 models. As can be seen by a comparison of the results for the pairs of models that differ only in terms of their treatment of $\gamma$ ($M_1$ versus $M_2$ etc.), the allowance for skewness does not provide an unambiguous improvement in within-sample fit. This result is consistent with the fact that the posterior estimates of $\gamma$ do not provide overwhelming evidence in favour of skewness. Both the BS model ($M_1$) and the skewed normal model with constant volatility ($M_2$) are inferior to all other models in terms of the RMSE, AIC and SIC criteria, as well as in terms of posterior probability.

### Table 6:

Measures of Predictive Performance

<table>
<thead>
<tr>
<th>Constant Volatility Models</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model RMSFE</td>
<td>MAPFE (%)</td>
<td>IQI (%)</td>
<td>60I (%)</td>
<td></td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.1644</td>
<td>47.31</td>
<td>11.30</td>
<td>40.87</td>
</tr>
<tr>
<td>$M_2$</td>
<td>0.1645</td>
<td>47.34</td>
<td>11.30</td>
<td>40.87</td>
</tr>
<tr>
<td>$M_3$</td>
<td>0.1618</td>
<td>46.34</td>
<td>14.78</td>
<td>42.61</td>
</tr>
<tr>
<td>$M_4$</td>
<td>0.1620</td>
<td>46.59</td>
<td>13.91</td>
<td>41.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time-Varying Volatility Models</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model RMSFE</td>
<td>MAPFE (%)</td>
<td>IQI (%)</td>
<td>60I (%)</td>
<td></td>
</tr>
<tr>
<td>$M_5$</td>
<td>0.1638</td>
<td>46.56</td>
<td>13.91</td>
<td>41.74</td>
</tr>
<tr>
<td>$M_6$</td>
<td>0.1646</td>
<td>46.70</td>
<td>13.04</td>
<td>40.87</td>
</tr>
<tr>
<td>$M_7$</td>
<td>0.1619</td>
<td>45.88</td>
<td>15.65</td>
<td>40.87</td>
</tr>
<tr>
<td>$M_8$</td>
<td>0.1630</td>
<td>46.03</td>
<td>14.78</td>
<td>40.87</td>
</tr>
</tbody>
</table>
4.3.3 Predictive Performance

Table 6 reports several measures of the predictive performance of the alternate models, based on the last 115 observations in the dataset as described in section 4.1. We include the proportion of observed prices contained within the interquartile interval (IQI) of the estimated predictive pdf (26) for each model, as well as the proportion of prices contained within the 60% interval (60I) that assigns 20% probability to each tail of the predictive, both expressed as a percentage of the number of out-of-sample observations. We also report the Root Mean Squared Forecast Error (RMSFE) and the Mean Absolute Percentage Forecast Error (MAPFE) for each model, calculated as

\[
RMSFE = \sqrt{\frac{1}{N_f} \sum_{f=1}^{N_f} \left( C_f - \left( \hat{\beta}_1 + \hat{\beta}_2 q_f \left( z_f, \hat{\theta}_k \right) \right)^2,\right.}
\]

and

\[
MAPFE = \frac{1}{N_f} \sum_{f=1}^{N_f} \left| \frac{C_f - \left( \hat{\beta}_1 + \hat{\beta}_2 q_f \left( z_f, \hat{\theta}_k \right) \right)}{C_f} \right| \times 100%.
\]

respectively, where \( \hat{\beta}_1, \hat{\beta}_2 \) and \( \hat{\theta}_k \) are as defined earlier, \( N_f = 115 \), and \( C_f \) and \( z_f \), \( f = 1, 2, \ldots, N_f \), are respectively the observed prices and the associated vector of contract characteristics in the out-of-sample period.

As with the corresponding within-sample fit measure, the MAPFE measure indicates that the time-varying volatility models perform better than the corresponding constant volatility models. The RMSFE measure however, does not confirm this result, in contrast to the within-sample case. Note that the out-of-sample performance of the models cannot be directly compared with the corresponding within-sample performance (via a comparison of the MAPFE/RMSFE and MAPE/RMSE statistics) as the characteristics of the hold-out sample (in terms of the moneyness/maturity characteristics of the option contracts) do not mimic those within-sample. The best performing models out-of-sample according to the MAPFE criterion, are the time-varying volatility models with a GST conditional distribution, \( M_7 \) and \( M_8 \). On the other hand, the constant volatility GST model, \( M_3 \), is the best performing model according to the RMSFE criterion, followed very closely by \( M_4 \) and \( M_7 \).

The IQI coverage statistics confirm the dominance of the time-varying volatility models, although the 60I coverage results are mixed. In both cases, the best perform-
ing model is deemed to be the one that produces an empirical coverage that is closest to the nominal level. According to the IQI criterion, $M_7$ performs best, with $M_8$ performing next best (equally with $M_3$). According to the 60I criterion however, the predictive performance of $M_7$ and $M_8$ is slightly inferior to that of the corresponding constant volatility GST models, $M_3$ and $M_4$ respectively. Model $M_3$ performs best overall in terms of the 60I coverage. Note that all models substantially underpredict according to the IQI and 60I statistics. The shape and location of the predictive densities reflect not only the specification of the underlying returns model, but also the specification of the regression structure in (12) and pricing error distribution in (13). Since these latter specifications are common to all models, the uniform result of underprediction may reflect a degree of misspecification in this component of the analysis.

4.3.4 Implied Volatility Graphs

For each model, implied volatilities are estimated from all option prices in the estimation sample associated with one particular time to maturity, namely 7 days. The value that is produced for implied volatility corresponds to $\exp(\delta_1)$, with $\delta_2$, $\nu$ and $\gamma$ set to their respective marginal posterior modes in the case of all models other than $M_1$. A quadratic function in the inverse of moneyness (denoted by $K/S$) is then fitted to the implied volatility data for each model. In each graph in Figure 1 the fitted curve associated with the BS model ($M_1$) is reproduced, with the curve associated with various of the other models superimposed. In this way, the impact on the shape of the implied volatility curve of modelling different distributional features, can be ascertained. If a model is adequate in capturing the distributional features implicit in the option price data, the smoothed graph of implied volatilities should be reasonably constant with respect to $K/S$. On the other hand, a pattern across $K/S$ suggests that option prices have factored in distributional assumptions that are not adequately captured by the model in question; for more details see Hull (2000, Chap. 17), Hafner & Herwartz (2001) and Lim et al. (2003).

As is clear from Figure 1, the skew that is typically associated with post-1987 BS implied volatilities for options on equities (see, for example, Corrado & Su, 1997, Bates, 2000 and Lim et al., 2003) is indeed a feature of the S&P200 option data, with the BS model essentially underpricing ITM options (low $K/S$) and overpricing OTM
options (high $K/S$). In the first panel of Figure 1, the implied volatility graph for the skewed normal model with constant volatility ($M_2$) is compared with the BS curve, illustrating that the modelling of skewness alone is insufficient in terms of producing a flat volatility graph. In contrast, in the adjacent panel, the specification of a skewed GST model for returns ($M_4$) produces an implied volatility graph that is virtually constant, at least for $K/S < 1$, with the graph for the corresponding symmetric model ($M_3$) being only slightly less constant over the same range. Interestingly,
as illustrated in the two lower panels of Figure 1, the incorporation of time-varying volatility does not aid in the production of a flat implied volatility graph. In summary then, according to this criterion the GST models with constant volatility, $M_3$ and $M_4$, are the best-performing models.

5 Concluding Remarks

The paper presents a general option pricing framework that accommodates the main empirical features of financial returns. In contrast with other attempts to generalize option pricing beyond the BS model, the approach adopted here has computational requirements equivalent to those of BS pricing, thereby rendering it a clear contender for use by practitioners. A Bayesian approach to conducting inference on the range of models accommodated within the general framework is outlined. When the methodology is applied to the prices of options on the S&P200 Index, there is clear evidence of option-implied time-varying volatility and excess kurtosis in Index returns, with slightly weaker evidence in favour of negative skewness. Whilst there is not complete consistency across all performance criteria, the results suggest that models that explicitly allow for these departures from the BS assumptions provide a better within-sample and out-of-sample fit to observed prices. In particular, those models that accommodate excess kurtosis in the distribution of returns feature prominently in terms of fit and predictive performance.

The practical import of the results presented here relates to the extent to which the better fit and predictive results associated with the non-BS models translate into better hedging and trading outcomes for participants in the Australian options market. Whilst several analyses assess the impact on hedging of the adoption of non-BS pricing (see, for example, Bakshi et al., 1997, Chernov & Ghysels, 2000 and Lim et al., 2003, amongst others), none of these analyses focus on Australian data. Furthermore, the emphasis of such work tends to be specifically on the impact on hedging errors of the adoption of more accurate pricing models. The broader impact of such pricing on the whole range of trading strategies adopted by market participants has yet to be explored, and represents a fruitful avenue for future research.
References


