ISSN 1440-771X

MONASH UNIVERSITY
AUSTRALIA

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Bayesian Estimation of a Stochastic Volatility Model Using Option and Spot Prices: Application of a Bivariate Kalman Filter

Catherine S. Forbes, Gael M. Martin and Jill Wright

Working Paper 17/2003

Bayesian Estimation of a Stochastic Volatility Model Using Option and Spot Prices: Application of a Bivariate Kalman Filter

Catherine S. Forbes, Gael M. Martin and Jill Wright

Department of Econometrics and Business Statistics, PO Box 11E,

Monash University, Vic. 3800 Australia. Email.

First Draft, February, 2002
(Monash University Working Paper 2/2002)
This Version, October, 2003

Abstract

In this paper Bayesian methods are applied to a stochastic volatility model using both the prices of the asset and the prices of options written on the asset. Posterior densities for all model parameters, latent volatilities and the market price of volatility risk are produced via a hybrid Markov Chain Monte Carlo sampling algorithm. Candidate draws for the unobserved volatilities are obtained by applying the Kalman filter and smoother to a linearization of a state-space representation of the model. The method is illustrated using the Heston (1993) stochastic volatility model applied to Australian News Corporation spot and option price data. Alternative models nested in the Heston framework are ranked via Bayes Factors and via fit, predictive and hedging performance.

Keywords: Option Pricing; Volatility Risk; Markov Chain Monte Carlo; Nonlinear State Space Model; Kalman Filter and Smoother.

JEL Classifications: C11, G13.

1 Introduction

In this paper we propose a Bayesian method for estimating a stochastic volatility model using both option prices and spot prices on the underlying asset. Posterior densities are produced for the parameters of the volatility model, for the latent volatilities and for the market price of volatility risk. The method involves augmenting the probability density function for a panel of option prices with the density function describing the bivariate process for the spot price and the volatility. Posterior results are produced via a hybrid Markov Chain Monte Carlo (MCMC) sampling algorithm. As part of this algorithm, candidate draws of the volatilities are obtained via the application of the Kalman filter and smoother to a linearization of a non-linear state-space representation of the model. Information from both the spot and option prices affects the draws via the specification of a bivariate measurement equation. In particular, a time series of implied volatilities produced via the Black and Scholes (1973) model is taken as a noisy measurement of the evolution of the assumed stochastic volatility process.

The method is illustrated using the Heston (1993) stochastic volatility model. In addition to demonstrating the production of marginal posteriors for the unknown elements of the Heston model, methods for comparing the Heston model with alternative models nested in the Heston framework are presented. These methods involve the construction of Bayes Factors as well as fit, predictive and hedging error densities. The methodology is applied to spot and option price data for News Corporation as observed over the period 1998 to 2001.

The outline of the paper is as follows. In Section 2, we briefly describe the Heston (1993) stochastic volatility model, making reference to other work in the literature that attempts to conduct inference on this model using observed option and spot prices. In Section 3, we describe our Bayesian inferential method, including the hybrid MCMC scheme that we adopt. We also outline the criteria used to rank nested versions of the Heston model, including the constant volatility Black-Scholes (BS) model. These criteria include posterior model probabilities, based in turn on Bayes Factors. The Bayes Factors are computed in a simple way, using the Savage-Dickey density ratio; see Koop and Potter (1999). Fit, predictive and hedging criteria to be used in model ranking are also detailed. This section also demonstrates how model averaging can be invoked to produce potentially more accurate predictions of future option prices in the case where market participants in fact price options via more than one distributional assumption. Section 4 includes a description of the News Corporation data to be used in the empirical demonstration of the method, followed by an outline of the numerical results. We provide some concluding comments in Section 5.

2 The Heston (1993) Stochastic Volatility Model

We begin by adopting the mean-reverting square root volatility process of Heston (1993). According to the Heston model, the risk-neutralized dynamics of the spot price and variance process respectively are:

$$dS(t) = rS(t)dt + \sqrt{v(t)}S(t)d\varepsilon_2(t) \tag{1}$$

and

$$dv(t) = \kappa[\theta - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_1(t), \tag{2}$$

where $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are correlated Weiner processes with correlation parameter ρ, r denotes the risk-free rate of interest and θ is the long-run mean of $v(t)$, to which $v(t)$ reverts at rate $\kappa > 0$. The *actual,* or *objective,* spot price and variance processes are given by:

$$dS(t) = \mu S(t)dt + \sqrt{v(t)}S(t)d\varepsilon_2(t) \tag{3}$$

and

$$dv(t) = \kappa^a[\theta^a - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_1(t), \tag{4}$$

where μ is the mean rate of return on the underlying asset. The parameters in (2) and (4) are related as follows:

$$\theta = \frac{\kappa^a\theta^a}{\kappa^a + \lambda}, \tag{5}$$

$$\kappa \quad = \quad \kappa^a + \lambda. \tag{6}$$

The volatility risk premium, $\lambda(S(t), v(t))$, is assumed to be proportional to $v(t)$, that is,

$$\lambda(S(t), v(t)) = \lambda v(t). \tag{7}$$

A negative value for λ and, hence, for the risk premium, is often observed empirically. With $\lambda < 0$, (2) implies slower reversion to a higher long-run mean than is implied by the actual variance process in (4).

Heston adopts standard arbitrage arguments to produce a closed form solution for the price of an option written on the underlying asset as:

$$q_H = S(t)P_1 - Ke^{-r\tau}P_2, \tag{8}$$

where q_H denotes the theoretical Heston option price, $S(t)$ denotes the current asset price, K denotes the strike, or exercise, price and $\tau = T - t$ denotes the time to maturity. The terms P_1 and P_2 are functions of the unknown parameters that characterize the risk-neutral volatility process, namely κ, θ, σ_v and the correlation parameter ρ, as well as being functions of the current latent variance, $v(t)$.[1] Alternatively, given (5) and (6), P_1 and P_2 can be viewed as functions of the parameters that characterize the *actual* volatility process, κ^a, θ^a, σ_v and ρ, the current latent variance, $v(t)$, and the risk premium parameter λ, as long as additional identifying information on the actual process is incorporated in the inferential procedure. We achieve this identification by augmenting the density function associated with the assumed generating process for the option prices with the density function that describes the dynamics of the spot prices and volatilities.

Guo (1998) and Chernov and Ghysels (2000) use classical methods to estimate the parameters of (2) using observed option price data. Guo uses time series data on returns to produce an estimate of θ^a directly, thereby enabling an estimate of λ to be backed out of the option prices, via the estimation of κ and θ. Parameter estimates are produced by minimizing the sum of squared differences between observed and theoretical option prices, with the minimization taken with respect to κ, θ, σ_v, ρ and the vector of latent variances, $v = \{v(t)\}$. Chernov and Ghysels use efficient method of moments to estimate the parameters of both the risk neutral and objective volatility processes. The data set constitutes both observations on the spot price $S(t)$ and a series of implied volatilities backed out, via the Black-Scholes (BS) model, from observed option prices. The form for the risk premium as given in (7) is generalized by the addition of a constant. Bates (1996 and 2002), Bakshi, Cao and Chen (1997) and Pan (2002) apply classical methods to an extension of the Heston model that accommodates random jumps in the asset price. Of these latter four works, only Pan uses both spot and options data and, hence, is able to produce inferences on both the underlying volatility process and the price of volatility risk.[2]

In this paper, a Bayesian approach to inference is adopted. As in Bates (1996 and 2000), we augment the density function for the option prices with the probability density function describing the actual volatility dynamics. However, in contrast to Bates, we also include the density function that describes the evolution of the spot price process, conditional on the volatility process. That is, we incorporate the bivariate spot price and volatility process as given in (3) and (4). This further augmentation of the data generating process enables identification of the parameters of both the actual volatility process and the risk-neutral process. Alternatively, it allows for the identification of the parameters of the actual process and the market price of volatility risk. We apply an MCMC sampling algorithm to produce marginal posterior densities for μ, κ^a, θ^a, σ_v, λ, ρ and the elements of v. The sampling algorithm is based on a hybrid of the Gibbs and Metropolis Hastings (MH) algorithms, with the MH subchains reweighting candidate draws according to the compatibility of the draws with the appropriate data set and associated data generating process. Specifically, both the option and spot price data provide information on v, κ^a, θ^a, σ_v and ρ. Since only the option price data reflects the market's attitude towards volatility risk, only that component of the data set provides information on λ. In contrast, since the options are priced assuming the risk-free rate of growth in the spot price, only the observed spot prices are relevant to inference on the actual mean rate of growth, μ.

There is some similarity between our proposed algorithm and the MCMC algorithm proposed independently by Eraker (2003), who estimates a modification of the Heston model using Bayesian methods.[3]However, in contrast to his work, we deal with the latent variances as (blocks of) the *vector*, v, rather than performing iterative simulation of each individual variance, $v(t)$, conditional on the remaining the variances. Also, in our method both the spot and option prices clearly impact on the simulation of the latent variances, via the bi-variate Kalman filter and smoother.[4]Further, we make explicit use of the approximate normality of the posterior of the risk premium parameter, λ, conditional on the vector of simulated variances, v, and on the values for the other parameters in the model. This approximate normality is exploited in the specification of the normal candidate density in the MH algorithm for λ. A similar approach is used in the specification of the (truncated) normal candidate density for ρ. Most importantly, the overall focus of our work is quite different from that of Eraker. Our aim is to demonstrate the application of a fully-fledged Bayesian approach to producing option-based inferences about stochastic volatility. That is, in addition to producing posterior point and interval estimates for the parameters of a particular stochastic volatility model, we demonstrate how to rank a set of alternative option pricing models, using Bayesian methods, as well as highlighting the relevance of Bayesian model averaging in an option pricing context. As the main aim of the paper is methodological, we choose to focus on the Heston model and three close variants, rather than expanding the model to accommodate the jump processes that may be needed in some empirical settings.[5]

3 The Bayesian Inferential Method

3.1 Specification of the Joint Posterior Density Function

Bayesian inferences about all unknown elements of the stochastic volatility model are to be produced in part from observed market option prices. For this to occur, option prices need to be assigned a particular distributional model. Letting C_i represent the ith observed market price of the call option and r_i, K_i, τ_i and S_i represent the observable factors that affect the ith option price, the option pricing model is specified as

$$C_i = \beta_0 + \beta_1 q_H(r_i, K_i, \tau_i, S_i, \phi) + u_i, \qquad i = 1, 2, \ldots, N, \tag{9}$$

where u_i is an unobservable pricing error, assumed to have a normal distribution with zero mean and variance, σ_u^2, N is the number of observed option prices and $q_H(r_i, K_i, \tau_i, S_i, \phi)$ is the ith theoretical option price as defined in (8). The vector $\phi = \{v, \omega^a, \lambda, \rho\}$ comprises all unobservable elements that characterize q_H, with the parameters of the objective volatility process in (4) grouped together in the vector $\omega^a = (\kappa^a, \theta_a, \sigma_v)'$. The index i indicates both variation over time and variation across option contracts at a given point in time.[6]

The vector of variances, v, has dimension equal to the number of distinct time periods, n say, in the pooled sample of option price data. Hence, $n < N$. In this paper, the dimension of v corresponds to the number of *days* over which the option price data are observed, with one variance per day, denoted by v_t, being estimated. Hence $v = (v_1, v_2, \ldots, v_n)'$. Each option price in the data set is assumed to be observed synchronously with a spot price S_i.[7]Hence, there are N observations (not necessarily distinct) on S_i. However, in specifying the joint density function for the spot prices and volatilities, as associated with a discretized version of (3) and (4), we use only the last spot price recorded on each day. Hence, the spot price process is assumed to describe movement in the volatilities and spot prices from day to day. We also use a single interest rate observation for each day, r_t, $t = 1, 2, \ldots, n$, where r_t denotes the 3 month bond rate on day t. It is notationally convenient, however, to continue to index all observable factors that influence the *ith* option price with i, and to group these factors together in a vector $z_i = (r_i, K_i, \tau_i, S_i)$, in which case, (9) becomes

$$C_i = \beta_0 + \beta_1 q_H(z_i, \phi) + u_i, \qquad i = 1, 2, \ldots, N. \tag{10}$$

The presence of u_i in (10) reflects the fact that the theoretical option model, $q_H(z_i, \phi)$, is only an approximation of the process that has lead to the determination of an observed option price. That is, u_i encompasses 'model error'. It may also encompass 'market error', in which an observed option price differs from its theoretical counterpart as the result of factors such as, for example, the non-synchronous recording of spot and option prices and transaction costs.[8]

The joint density function for the vector of option prices $c = (C_1, C_2, \ldots, C_N)'$, conditional on the known vector $z = (z_1, z_2, \ldots z_N)'$ and on the unknown ϕ, is thus given by

$$p(c|z,\phi, \beta, \sigma_u) = (2\pi\sigma_u^2)^{-N/2} \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma_u^2}[C_i - [\beta_0 + \beta_1 q_H(z_i, \phi)]]^2\right). \tag{11}$$

As already noted, in order to produce simultaneous inference about the parameters of both the risk-neutral and objective processes or, equivalently, about the parameters of the objective process and the market price of volatility risk, information about the way in which the observed spot prices have evolved needs to be incorporated in the inferential procedure. We incorporate this information by augmenting the data generating process in (11) with a discretized bivariate spot price and volatility process in (3) and (4). This augmentation also serves to identify the process to which the estimated volatilities must adhere.[9]

The vector of (closing) spot prices associated with the n days in the sample is defined as $s = (S_1, S_2, \ldots, S_n)'$. Suppressing the dependence of $p(c|z,\phi)$ on all elements of z other than s (including all other intraday synchronous spot prices that are used to calculate $q_H(z_i, \phi)$ for each i) and defining the vector $\delta = \{\phi, \mu, \beta, \sigma_u\}$ as the full set of unknowns in the problem, with $\beta = (\beta_0, \beta_1)'$, the joint posterior density for δ is thus specified as

$$\begin{aligned} p(\delta|c,s) &\propto p(c|s,v,\omega^a,\lambda,\rho,\beta,\sigma_u) \times p(s,v|\omega^a,\rho,\mu) \times p(\omega^a,\rho,\lambda,\mu,\beta,\sigma_u) \\ &\propto p(c|s,v,\omega^a,\lambda,\rho,\beta,\sigma_u) \times p(s|v,\omega^a,\rho,\mu) \times p(v|\omega^a) \\ &\quad \times p(\omega^a,\rho,\lambda,\mu,\beta,\sigma_u). \end{aligned} \tag{12}$$

The posterior density in (12) contains all information, both sample-based and *a priori*, regarding the elements of δ. We further specify that

$$p(\omega^a,\rho,\lambda,\mu,\beta,\sigma_u) = p(\omega^a) \times p(\rho) \times p(\lambda) \times p(\mu) \times p(\beta) \times p(\sigma_u). \tag{13}$$

That is, it is assumed that the set of parameters that characterize the volatility process, namely ω^a, is *a priori* independent of the mean rate of return on the underlying asset, μ, as well as being independent of both λ and ρ. The latter two parameters are also assumed to be *a priori* independent of one another.[10]The regression parameters β and σ_u are assumed to be *a priori* independent both of each other and of all other parameters in the model.

3.2 The Hybrid Gibbs-MH Algorithm

Due to the large number of unknowns in the model and the manner in which they are related, computation of the joint posterior distribution and marginal posterior distributions is not possible analytically, and an MCMC algorithm has been developed. To implement the MCMC algorithm, the parameters in δ are 'blocked' into eight groups[11]as follows: v, κ^a, θ_a, σ_v, λ, ρ, μ and (β, σ_u). Starting values are chosen, and a Gibbs-based MCMC algorithm is then applied to produce successive draws of the unknowns via the respective conditional posteriors:

1. $p(v|\omega_a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$
2. $p(\kappa^a|v, \theta^a, \sigma_v, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$
3. $p(\theta^a|v, \kappa^a, \sigma_v, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$
4. $p(\sigma_v|v, \kappa^a, \theta^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$
5. $p(\lambda|v, \omega_a, \rho, \beta, \sigma_u, c, s)$
6. $p(\rho|v, \omega_a, \lambda, \mu, \beta, \sigma_u, c, s)$
7. $p(\mu|v, \omega_a, \rho, s)$
8. $p(\beta, \sigma_u|v, \omega_a, \lambda, \rho, c, s)$

In the description of each conditional, we make explicit the relevant conditioning elements. For example, the conditionals of λ and (β, σ_u) do not depend on μ and the conditional of μ does not depend on any aspect of the model that relates to the option prices, namely (β, σ_u), λ and c. All of the conditionals, apart from those of μ and (β, σ_u), are nonstandard, with MH subchains being applied to produce draws. Given the satisfaction of the relevant regularity conditions (see Tierney, 1994), the draws obtained from the hybrid Gibbs-MH algorithm converge in distribution to a sample from the full joint posterior distribution. We consider the eight conditionals in order.

3.2.1 $p(v|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$

Conditional on values for ω^a, λ, ρ, μ , β and σ_u the posterior density for v is given by

$$p(v|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s) \propto p(c|s,v, \omega^a, \lambda, \rho, \beta, \sigma_u) \times p(s|v, \omega^a, \rho, \mu) \times p(v|\omega^a). \quad (14)$$

From (14) it follows that the ordinate of the conditional posterior for v, at some value, v^* say, is equal, up to a scale factor, to the product of the ordinate of the joint density function for the option prices, conditioned on v^*, the ordinate of the joint density function for the spot prices, conditioned on v^*, and the ordinate of the joint density function for v, evaluated at v^* (with all density functions also dependent on the conditioning values for the relevant parameters).

The MH algorithm for simulating from the conditional posterior for the stochastic variance vector involves the specification of a candidate model for v. The resulting candidate posterior probability distribution has a joint density function denoted by $p_c(v|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$. The form of the candidate model suggested here is based upon a linearization of a state-space representation of the Heston model. This representation is further augmented by the construction of a second observation equation which specifies that the BS implied variances, calculated from the observed option prices, are noisy measurements of the elements of the vector of true stochastic variances, v. In order to produce a vector of implied BS variances that matches the dimension of v, the BS variances are averaged over the day, with a vector of n BS variances produced as a result.

The form of $p(v|\omega^a)$ in (14) is based on a Euler discretization of (4):

$$\begin{aligned} v_t &= v_{t-1} + \kappa^a(\theta^a - v_{t-1})\Delta t + \sigma_v\sqrt{v_{t-1}}\sqrt{\Delta t}\varepsilon_{1t} \\ &= \kappa^a\theta^a\Delta t + (1-\kappa^a\Delta t)v_{t-1} + \sigma_v\sqrt{v_{t-1}}\sqrt{\Delta t}\varepsilon_{1t} \end{aligned} \tag{15}$$

$$\varepsilon_{1t} \sim N(0,1); \quad 2, 3, \ldots n,$$

with the initial value set equal to the long-run mean, θ^a. By convention both the variances and the parameters of the volatility model enter the theoretical option price formula, $q_H(z_i, \phi)$, in annualized form. As (15) describes day-to-day movements in the annualized variance, we set $\Delta t = 1/252$ (years), assuming 252 trading days in the year; see also Chernov and Ghysels (2000).[12]

The form of $p(s|v, \omega^a, \rho, \mu)$ in (14) is determined via an Euler discretization of (3), namely

$$\ln S_t = \ln S_{t-1} + (\mu - 0.5v_{t-1})\Delta t + \sqrt{v_{t-1}}\sqrt{\Delta t}\varepsilon_{2t}, \tag{16}$$

$$\varepsilon_{2t} \sim N(0,1); \quad 2, 3, \ldots n,$$

conditional on a specified value for $\ln S_1$. From (16) it follows that

$$\begin{aligned} p(\ln s|v, \omega^a, \rho, \mu) &= (2\pi)^{-(n-1)/2}[(1-\rho^2)\Delta t]^{-(n-1)/2} \times \prod_{t=2}^{n} \frac{1}{\sqrt{v_{t-1}}} \\ & \exp\left\{-1/[2(1-\rho^2)\Delta t] \sum_{t=2}^{n} (\frac{\ln S_t - \mu_{\ln S_t.v_t}}{\sqrt{v_{t-1}}})^2\right\} \end{aligned} \tag{17}$$

where

$$\mu_{\ln S_t.v_t} = (\ln S_{t-1} + [\mu - 0.5v_{t-1}]\Delta t) + \frac{\rho}{\sigma_v}(v_t - [\theta^a \kappa^a \Delta t + (1 - \kappa^a \Delta t)v_{t-1}]). \tag{18}$$

Equation (15) describes the evolution of the variances, whereas the expressions in (17) and (18) together describe the probabilities associated with the observation of the logarithm of the current spot price, $\ln S_t$, conditional upon the value of the most recent spot price, S_{t-1}, the values of the current and most recent variances, v_t and v_{t-1}, respectively, and the unknown parameters, ω^a, λ, ρ, and μ.

The discretized model in (15) and (16) can be written as a nonlinear state space model. The measurement of the spot return, $\Delta \ln S_t$, depends upon the unobservable volatilities v_t and v_{t-1}. To introduce information from the option prices, we further augment this model by specifying that the implied BS variance for day t, v_t^{BS}, is an independent, noisy measurement of the true stochastic variance, v_t. The variance of this measurement is fixed, and denoted by σ^2_{imp}. A state space representation of the augmented model is therefore given by the following two bivariate equations

$$\begin{bmatrix} \Delta \ln S_t \\ v_t^{BS} \end{bmatrix} = \begin{bmatrix} \frac{\rho}{\sigma_v} & -\left(\frac{\Delta t}{2} + \frac{\rho(1-\kappa^a \Delta t)}{\sigma_v}\right) \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} + \begin{bmatrix} \mu \Delta t - \frac{\rho \theta^a \kappa^a \Delta t}{\sigma_v} \\ 0 \end{bmatrix} + \begin{bmatrix} \sqrt{v_{t-1}\Delta t\,(1-\rho^2)}\varepsilon_{2t} \\ \sigma_{imp}\varepsilon_{3t} \end{bmatrix} \tag{19}$$

and

$$\begin{bmatrix} v_t \\ v_{t-1} \end{bmatrix} = \begin{bmatrix} 1 - \kappa^a \Delta t & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v_{t-1} \\ v_{t-2} \end{bmatrix} + \begin{bmatrix} \theta^a \kappa^a \Delta t \\ 0 \end{bmatrix} + \begin{bmatrix} \sigma_v \sqrt{v_{t-1}\Delta t}\varepsilon_{1t} \\ 0 \end{bmatrix}. \tag{20}$$

To eradicate the nonlinear dependence of both the return, $\Delta \ln S_t$, and the current stochastic variance, v_t, on the previous stochastic variance, v_{t-1}, we replace v_{t-1} in the error specification with its long-run mean, θ^a. It is this linearized approximation to (19) and (20) that is used in the MH algorithm, with steps described as follows:

Step 1 For given values of the parameters $\omega^a, \lambda, \rho, \mu$, simulate a vector of variances, v^*, from the candidate density, $p_c(v|\omega^a, \lambda, \rho, \mu, c, s)$, obtained by running a Kalman filter on the linearized state space model and subsequently drawing elements of v^* using a backwards simulation smoother as described, for example, in de Jong and Shephard (1995).

Step 2 Select the simulated vector value, v^*, as a drawing from $p(v|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$ with probability

$$\begin{aligned} \psi &= \min\left\{\frac{p(v^*|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)}{p_c(v^*|\omega^a, \lambda, \rho, \mu, c, s)} / \frac{p(v^s|\omega^a, \lambda, \rho, \mu, \beta, \sigma_u, c, s)}{p_c(v^s|\omega^a, \lambda, \rho, \mu, c, s)}, 1\right\} \\ &= \min\{\frac{p\left(c|s, v^*, \omega^a, \lambda, \rho, \beta, \sigma_u\right) \times p(s|v^*, \omega^a, \rho, \mu) \times p(v^*|\omega^a)}{p_c(v^*|\omega^a, \lambda, \rho, \mu, c, s)} \\ &\quad / \frac{p\left(c|s, v^s, \omega^a, \lambda, \rho, \beta, \sigma_u\right) \times p(s|v^s, \omega^a, \rho, \mu) \times p(v^s|\omega^a)}{p_c(v^s|\omega^a, \lambda, \rho, \mu, c, s)}, 1\}, \end{aligned} \quad (21)$$

where v^s indicates a starting value for the MH subchain. Note that the evaluation of the candidate density at the previously generated volatility vector, v^s, requires a rerunning of the Kalman smoothing algorithm using the v^s values to obtain the updated means and variances.[13]Note also that, since the Jacobian of the transformation from $\ln S_t$ to S_t is not a function of any of the elements of δ, it follows that

$$\frac{p(\ln s|v^*, \omega^a, \rho, \mu)}{p(\ln s|v^s, \omega^a, \rho, \mu)} = \frac{p(s|v^*, \omega^a, \rho, \mu)}{p(s|v^s, \omega^a, \rho, \mu)}$$

and that the joint density in (17) can be used in the MH selection algorithm. In fact, this density is used in the algorithms for all blocks of δ. However, for notational simplicity, we continue to refer to the density $p(s|v, \omega^a, \rho, \mu)$ in the description of these algorithms.

3.2.2 $p(\omega^a|v, \lambda, \rho, \mu, \beta, \sigma_u, c, s)$

Conditional on values for v, λ, ρ, μ, β, and σ_u, the posterior density for $\omega^a = (\kappa^a, \theta_a, \sigma_v)'$ is given by

$$\begin{aligned} p(\omega^a|v, \lambda, \rho, \mu, \beta, \sigma_u, c, s) &\propto p\left(c|s, v, \omega^a, \lambda, \rho, \beta, \sigma_u\right) \times p(s|v, \omega^a, \rho, \mu) \times p(v|\omega^a) \times p(\omega^a) \\ &\propto p\left(c|s, v, \omega^a, \lambda, \rho, \beta, \sigma_u\right) \times p(s|v, \omega^a, \rho, \mu) \times p(\omega^a|v). \end{aligned} \quad (22)$$

From (22) it follows that the ordinate of the conditional posterior for ω^a, at some vector value ω^{a*} say, is equal, up to a scale factor, to the product of the ordinate of the joint density function for the observed option prices, conditioned on ω^{a*}, the ordinate of the joint density function for the observed spot prices, conditioned on ω^{a*}, and the ordinate of the joint density function for ω^a, evaluated at ω^{a*}, given the conditioning value v. The latter density function, $p(\omega^a|v)$, is like a posterior density function for ω^a given 'data' v and, hence, reflects both the assumed generating process for v, as specified in (4), as well as the prior on ω^a. Again using the Euler discretization for (4) as given in (15), the form of $p(\omega^a|v)$ is as follows. Defining

$$\zeta = 1 - \kappa^a \Delta t,$$

(15) can be rewritten as:

$$\begin{aligned} v_t &= \kappa^a \theta^a \Delta t + (1 - \kappa^a \Delta t) v_{t-1} + \sigma_v \sqrt{v_{t-1}} \sqrt{\Delta t} \varepsilon_{1t} \\ &= \theta^a (1 - \zeta) + \zeta v_{t-1} + \sigma_v \sqrt{v_{t-1}} \sqrt{\Delta t} \varepsilon_{1t}. \end{aligned} \tag{23}$$

Further defining

$$\zeta(L) = 1 - \zeta L, \tag{24}$$

$$y_t = \frac{\zeta(L) v_t}{\sqrt{v_{t-1}} \sqrt{\Delta t}}, \tag{25}$$

$$x_t = \frac{1 - \zeta}{\sqrt{v_{t-1}} \sqrt{\Delta t}} \tag{26}$$

and

$$e_t = \sigma_v \varepsilon_{1t},$$

where L is the lag operator with respect to time period Δt, (23) can be written as

$$\begin{aligned} y_t &= \theta^a x_t + e_t, \\ e_t &\sim N(0, \sigma_v^2), \\ t &= 2, 3, \ldots, n. \end{aligned} \tag{27}$$

Given (27), the form of $p(\omega^a|v)$ is thus

$$p(\omega^a|v) \propto \sigma_v^{-n} \exp\left\{ \frac{-1}{2\sigma_v^2} \sum_{t=1}^{n} (y_t - \theta^a x_t)^2 \right\} \times p(\omega^a), \tag{28}$$

where

$$
\begin{aligned}
p(\omega^a) &= p(\kappa^a)p(\theta^a)p(\sigma_v) \\
&\propto 1_{(\kappa^a>0)} \times 1_{(\theta^a>0)} \times 1_{(2\kappa^a\theta^a>\sigma_v^2)} \times \frac{1}{\sigma_v}. \qquad (29)
\end{aligned}
$$

With reference to (29) the first two indicator functions, $1_{(\kappa^a>0)}$ and $1_{(\theta^a>0)}$, restrict κ^a and θ^a respectively to the positive region[14]and the third indicator function, $1_{(2\kappa^a\theta^a>\sigma_v^2)}$, ensures that the variances associated with (4) are always positive.

The elements of ω^a are drawn one at a time, with each component of $p(\omega^a|v,\lambda,\rho,\mu,\beta,\sigma_u,c,s)$ in (22) viewed as a function of the relevant parameter conditional on the other two parameters (and the remaining conditional elements). For drawing κ^a, the following random walk MH algorithm is used:

Step 1 Generate κ^a from a normal candidate distribution, with mean equal to the previous draw in the outer Gibbs chain, denoted by $\kappa^{a(i-1)}$, and variance, σ_κ^2, tuned in a preliminary algorithm to produce an acceptance rate of between approximately 20% and 70%.

Step 2 Select the drawn value, κ^{a*}, as a drawing from $p(\kappa^a|\theta^a,\sigma_v,v,\lambda,\rho,\mu,\beta,\sigma_u,c,s)$ with probability

$$
\begin{aligned}
\psi &= \min\left\{p(\kappa^{a*}|\theta^a,\sigma_v,v,\lambda,\rho,\mu,\beta,\sigma_u,c,s)/p(\kappa^{as}|\theta^a,\sigma_v,v,\lambda,\rho,\mu,\beta,\sigma_u,c,s),1\right\} \\
&= \min\{\frac{p\left(c|s,v,\kappa^{a*},\theta^a,\sigma_v,\lambda,\rho,\beta,\sigma_u\right)\times p(s|v,\kappa^{a*},\theta^a,\sigma_v,\rho,\mu)\times p(\kappa^{a*}|\theta^a,\sigma_v,v)}{p\left(c|s,v,\kappa^{as},\theta^a,\sigma_v,\lambda,\rho,\beta,\sigma_u\right)\times p(s|v,\kappa^{as},\theta^a,\sigma_v,\rho,\mu)\times p(\kappa^{as}|\theta^a,\sigma_v,v)},1\},
\end{aligned}
$$

where κ^{as} indicates a starting value for the MH subchain and $p(\kappa^a|\theta^a,\sigma_v,v)$ denotes the density for κ^a implied by (28). That is, the MH subchain involves assessing the ratio of the relative likelihoods of c given simulated and previous values for the mean reversion parameter of the variance process, multiplied by the ratio of corresponding likelihoods for the spot price data and the ratio of the (conditional) 'posterior' ordinates for κ^a, given the 'data' v.

The parameters θ^a and σ_v are drawn in an analogous fashion to κ^a.

3.2.3 $p(\lambda|v,\omega^a,\rho,\beta,\sigma_u,c,s)$

The conditional posterior for λ is given by

$$p(\lambda|v,\omega^a,\rho,\beta,\sigma_u,c,s) \propto p\left(c|s,v,\omega^a,\lambda,\rho,\beta,\sigma_u\right) \times p(\lambda). \tag{30}$$

To sample from (30) we adopt an MH algorithm, based on a normal candidate density. A normal candidate is adopted due to the accuracy with which it has been found to approximate the actual conditional for λ (based on a uniform prior for λ), in preliminary investigations. This empirical regularity implies that the theoretical option price, $q_H(z_i,\phi)$, conditional on given values for v, ω^a and ρ, is approximately linear in λ. We use a Taylor Series expansion of $q_H(z_i,\lambda|v,\omega^a,\rho)$ around $\lambda=\lambda^{\#}$ to represent this linear relationship as follows,

$$q_H(z_i,\lambda|v,\omega^a,\rho) \approx q_H(z_i,\lambda^{\#}|v,\omega^a,\rho) + q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho)(\lambda-\lambda^{\#}), \tag{31}$$

where $q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho)$ denotes the first derivative of $q_H(z_i,\lambda|v,\omega^a,\rho)$ with respect to λ, evaluated at $\lambda^{\#}$. This first derivative is, in turn, approximated as

$$q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho) \approx \frac{q_H(z_i,(\lambda^{\#}+h)|v,\omega^a,\rho) - q_H(z_i,\lambda^{\#}|v,\omega^a,\rho)}{h} \tag{32}$$

for h small. Substitution of (31) and (32) for $q_H(z_i,\lambda|v,\omega^a,\rho)$ in the expression for $p\left(c|s,v,\omega^a,\lambda,\rho,\beta,\sigma_u\right)$ in (30) produces a conditional candidate density, $p_c(\lambda|v,\omega^a,\rho,\beta,\sigma_u,c,s)$, of the form

$$p_c(\lambda|v,\omega^a,\rho,\beta,\sigma_u,c,s) \propto \exp\left(\frac{-\beta_1^2\sum_{i=1}^N(q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho))^2}{2\sigma_u^2}(\lambda-\widehat{\lambda})^2\right) \times p(\lambda),$$

where

$$\widehat{\lambda} = \frac{\sum_{i=1}^N(C_i-\beta_0-\beta_1[q_H(z_i,\lambda^{\#}|v,\omega^a,\rho) - q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho)\lambda^{\#}]q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho)}{\beta_1\sum_{i=1}^N(q_H^{'}(z_i,\lambda^{\#}|v,\omega^a,\rho))^2}. \tag{33}$$

Adopting a normal prior for λ, with mean $\overline{\lambda}$ and variance $\overline{var}(\lambda)$[15], the normal candidate density for λ is given by

$$p_c(\lambda|v,\omega^a,\rho,\beta,\sigma_u,c,s) \propto \exp\left(\frac{-1}{2\overline{\overline{var}}(\lambda)}(\lambda-\overline{\overline{\lambda}})^2\right), \tag{34}$$

with

$$\begin{aligned}\overline{\overline{\lambda}} &= \left[\frac{\beta_1^2 \sum_{i=1}^N (q_H^{'}(z_i, \lambda^{\#}|v, \omega^a, \rho))^2}{\sigma_u^2}\widehat{\lambda} + \frac{1}{\overline{var}(\lambda)}\overline{\lambda}\right] \\ &\quad / \left[\frac{\beta_1^2 \sum_{i=1}^N (q_H^{'}(z_i, \lambda^{\#}|v, \omega^a, \rho))^2}{\sigma_u^2} + \frac{1}{\overline{var}(\lambda)}\right]\end{aligned}$$

and

$$\overline{\overline{var}}(\lambda) = 1/\left[\frac{\beta_1^2 \sum_{i=1}^N (q_H^{'}(z_i, \lambda^{\#}|v, \omega^a, \rho))^2}{\sigma_u^2} + \frac{1}{\overline{var}(\lambda)}\right].$$

In the numerical application, we specify $\lambda^{\#}$ to be the value of λ produced in the previous iteration of the outer Gibbs chain.

In the usual way, a candidate value, λ^* is drawn from $p_c(\lambda|v, \omega^a, \rho, \mu, \beta, \sigma_u, c, s)$ and chosen with probability,

$$\psi = \min\left\{\frac{p(\lambda^*|v, \omega^a, \rho, \beta, \sigma_u, c, s)}{p_c(\lambda^*|v, \omega^a, \rho, \beta, \sigma_u, c, s)} / \frac{p(\lambda^s|v, \omega^a, \rho, \beta, \sigma_u, c, s)}{p_c(\lambda^s|v, \omega^a, \rho, \beta, \sigma_u, c, s)}, 1\right\},$$

where λ^s indicates a starting value for the MH subchain.

3.2.4 $p(\rho|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)$

The treatment of the parameter ρ is analogous to the treatment of λ, except for the fact that the candidate density, $p_c(\rho|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)$, is the product of two normal approximations, to $p\,(c|s, v, \omega^a, \lambda, \rho, \beta, \sigma_u)$ and $p(s|v, \omega^a, \rho, \mu)$ respectively, and a normal prior for ρ. Combining the prior density with the normal approximation to $p\,(c|s, v, \omega^a, \lambda, \rho, \beta, \sigma_u)$, the first component of the candidate is defined as a normal density for ρ, with mean and variance given respectively by

$$\begin{aligned}\overline{\overline{\rho}} &= \left[\frac{\beta_1^2 \sum_{i=1}^N (q_H^{'}(z_i, \rho^{\#}|v, \omega^a, \lambda))^2}{\sigma_u^2}\widehat{\rho} + \frac{1}{\overline{var}(\rho)}\overline{\rho}\right] \\ &\quad / \left[\frac{\beta_1^2 \sum_{i=1}^N (q_H^{'}(z_i, \rho^{\#}|v, \omega^a, \lambda))^2}{\sigma_u^2} + \frac{1}{\overline{var}(\rho)}\right]\end{aligned}$$

and

$$\overline{\overline{var}}(\rho) \quad = \quad 1/\left[\frac{\beta_1^2 \sum_{i=1}^{N}(q_H^{'}(z_i,\rho^{\#}|v,\omega^a,\lambda))^2}{\sigma_u^2} + \frac{1}{\overline{var}(\rho)}\right],$$

where $q_H^{'}(z_i,\rho^{\#}|v,\omega^a,\lambda)$ denotes the first derivative of $q_H(z_i,\rho|v,\omega^a,\lambda)$, evaluated at $\rho = \rho^{\#}$, $\overline{\rho}$ and $\overline{var}(\rho)$ are respectively the mean and variance of the prior normal distribution for ρ and

$$\widehat{\rho} = \frac{\sum_{i=1}^{N}(C_i - \beta_0 - \beta_1[q_H(z_i,\rho^{\#}|v,\omega^a,\lambda) - q_H^{'}(z_i,\rho^{\#}|v,\omega^a,\lambda)\rho^{\#}]q_H^{'}(z_i,\rho^{\#}|v,\omega^a,\lambda)}{\beta_1 \sum_{i=1}^{N}(q_H^{'}(z_i,\rho^{\#}|v,\omega^a,\lambda))^2.}$$

The second component of the candidate, based on a normal approximation to $p(s|v,\omega^a,\rho,\mu)$, is defined as a normal density for ρ, with mean and variance given respectively by

$$\widehat{\rho}_s = \left(\frac{\sum_{t=2}^{n} y_t^{(\rho)} x_t^{(\rho)}}{\sum_{t=2}^{n}\left(x_t^{(\rho)}\right)^2}\right)$$

and

$$\sigma_{\rho_s}^2 = \left(\frac{\Delta t}{\sum_{t=2}^{n}\left(x_t^{(\rho)}\right)^2}\right)$$

where

$$y_t^{(\rho)} \quad = \quad \frac{(\ln S_t - \ln S_{t-1} - (\mu - 0.5 v_{t-1})\Delta t)}{\sqrt{v_{t-1}}}$$

and

$$x_t^{(\rho)} \quad = \quad \left(\frac{(v_t - \theta^a \kappa^a \Delta t - (1 - \kappa^a \Delta t)v_{t-1})}{\sqrt{v_{t-1}}}\right) \times \frac{1}{\sigma_v}.$$

The product of these two normal components is used to produce a candidate draw for ρ, ρ^*, which is in turn selected as a draw from the conditional posterior for ρ,

$p(\rho|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)$, with probability

$$\psi = \min\left\{ \frac{p(\rho^*|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)}{p_c(\rho^*|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)} / \frac{p(\rho^s|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)}{p_c(\rho^s|v, \omega^a, \lambda, \mu, \beta, \sigma_u, c, s)}, 1 \right\},$$

where ρ^s indicates a starting value for the MH subchain. Since the candidate density needs to reflect the truncation of the actual conditional for ρ at ± 1, the draws from the candidate density are discarded if they fall beyond these bounds.

3.2.5 $p(\mu|v, \omega^a, \rho, s)$

Using the expressions in (17) and (18) and adopting a uniform prior for μ, it follows that $p(\mu|v, \omega^a, \rho, s)$ is normal with mean

$$\widehat{\mu} = \left(\frac{\sum_{t=2}^{n} y_t^{(\mu)} x_t^{(\mu)}}{\sum_{t=2}^{n} \left(x_t^{(\mu)} \right)^2} \right) \times \frac{1}{\Delta t}$$

and variance

$$\sigma_\mu^2 = \left(\frac{(1-\rho^2)}{\sum_{t=2}^{n} \left(x_t^{(\mu)} \right)^2} \right) \times \frac{1}{\Delta t},$$

with

$$y_t^{(\mu)} = \frac{\left(\ln S_t - \ln S_{t-1} + 0.5 v_{t-1} \Delta t - \frac{\rho}{\sigma_v} (v_t - [\theta^a \kappa^a \Delta t + (1 - \kappa^a \Delta t) v_{t-1}] \right)}{\sqrt{v_{t-1}}}$$

and

$$x_t^{(\mu)} = \frac{1}{\sqrt{v_{t-1}}}.$$

Simulated values of μ are thus readily obtainable via the generation of normal random variates.

3.2.6 $p(\beta, \sigma_u | v, \omega^a, \lambda, \rho, c, s)$

Using the expression for the joint density of the vector of option prices in (11) and adopting the following standard noninformative prior for (β, σ_u),

$$\begin{aligned} p(\beta, \sigma_u) &= p(\beta) \times p(\sigma_u) \\ &\propto \frac{1}{\sigma_u}, \end{aligned}$$

the joint conditional density for (β, σ_u) has the following normal-gamma form,

$$p(\beta, \sigma_u | v, \omega^a, \lambda, \rho, c, s) = p(\beta | \sigma_u, v, \omega^a, \lambda, \rho, c, s) p(\sigma_u | v, \omega^a, \lambda, \rho, c, s). \tag{35}$$

The density $p(\beta | \sigma_u, v, \omega^a, \lambda, \rho, c, s)$ in (35) is normal with mean

$$\widehat{\beta} = (X'X)^{-1} X'c \tag{36}$$

and variance

$$var(\beta) = \sigma_u^2 (X'X)^{-1}, \tag{37}$$

where $X = (\iota, q_H)$, with ι a $(N \times 1)$ vector of ones and q_H denoting the $(N \times 1)$ vector of theoretical option prices with ith element $q_H(z_i, \phi)$. The density $p(\sigma_u | v, \omega^a, \lambda, \rho, c, s)$ in (35) is inverted gamma with degrees of freedom $\nu_{\sigma_u} = (N - 2)$ and parameter $s_{\sigma_u} = \sqrt{(c - X\widehat{\beta})'(c - X\widehat{\beta}) / \nu_{\sigma_u}}$. Draws of (β, σ_u) can be obtained from (35) using standard simulation algorithms.

Implementation of the hybrid MCMC scheme requires only one draw for each of the MH subchains (see Chib and Greenberg, 1996, on this point). All posterior quantities of interest are to be calculated from the full set of MCMC iterates, excluding those in the burn-in part of the chain. Marginal posteriors are to be estimated from the simulated values for each parameter of interest using kernel smoothing.

3.3 Model Ranking and Model Averaging

The Heston (1993) model nests three alternative models for volatility, associated respectively with: $\lambda = 0$, $\rho = 0$ and $\sigma_v = 0$. Setting $\lambda = 0$ is equivalent to imposing the assumption that volatility risk is not priced. This assumption is invoked in the early stochastic volatility analysis of Hull and White (1987) for computational convenience. It has however been challenged by more recent work, in which estimates of λ that differ significantly from zero have been reported (see Guo, 1998 and Eraker, 2003, amongst others). An assessment of this restricted model thus amounts to an assessment of the attitude to volatility risk that is implicit in option prices.

The model obtained by setting $\rho = 0$ implies a lack of the so-called "leverage effect" (associated with $\rho < 0$), whereby negative returns are accompanied by an increase in volatility. Since this effect corresponds, in turn, to the empirical characteristic of negative skewness in returns, an assessment of this restricted model corresponds to an assessment of whether or not returns are skewed and/or option prices have factored in skewed returns.

Finally, the restriction $\sigma_v = 0$ implies constant volatility. Given the assumption of normal returns, this restriction equates to the assumption of BS option pricing. An assessment of the empirical validity of this restriction thus amounts to an assessment of the validity of the BS model.

We refer to the alternative models corresponding to the restrictions $\lambda = 0$, $\rho = 0$ and $\sigma_v = 0$ as respectively M_2, M_3 and M_4, and to the full Heston model as M_1. In this section, several criteria that are used to rank these alternative models in the empirical section are described. These criteria are used to supplement the results obtained by simply estimating the full model, M_1, and testing the restrictions via the construction of interval estimates for each of the relevant parameters. The concept of averaging across the alternative models, in particular with a view to improving predictive performance, is also discussed. In what follows, we refer to the vectors of unobservables associated with the four models, M_1, M_2, M_3 and M_4, as δ_1, δ_2, δ_3 and δ_4 respectively. These vectors are in turn defined as follows:

1. $\delta_1 = \{v, \omega^a, \lambda, \rho, \mu, \beta, \sigma_u\}$
2. $\delta_2 = \{v, \omega^a, \rho, \mu, \beta, \sigma_u\}$
3. $\delta_3 = \{v, \omega^a, \lambda, \mu, \beta, \sigma_u\}$
4. $\delta_4 = \{\theta^a, \mu, \beta, \sigma_u\}$

Due to the nested structure of these models, we can also re-express δ_1 as, alternatively, $\delta_1 = \{\delta_2, \lambda\}$, $\delta_1 = \{\delta_3, \rho\}$ or $\delta_1 = \{\delta_4, v, \kappa^a, \sigma_v, \lambda, \rho\}$.

3.3.1 Bayes Factors using the Savage-Dickey Density Ratio

In the Bayesian framework, the four alternative models, M_1, M_2, M_3 and M_4, can in principle be ranked according to the magnitude of their respective posterior probabilities. In the present context, the data on which posterior inference is based comprises the vector of option prices, c, and the vector of spot prices, s. Hence, the ranking occurs via the model probabilities, $P(M_1|c, s)$, $P(M_2|c, s)$, $P(M_3|c, s)$ and $P(M_4|c, s)$. These probabilities can, in turn, be derived via the set of posterior odds ratios for each model, relative to the reference model, M_1, subject to the restriction that the posterior probabilities add to one. Given equal prior odds for all models, the posterior odds ratio for M_k versus M_1 reduces to the Bayes Factor for M_k versus M_1,

$$BF_{k1} = \frac{p(c, s|M_k)}{p(c, s|M_1)}; \quad k = 2, 3, 4, \tag{38}$$

where $p(c, s|M_k)$ denotes the marginal likelihood for model M_k and is defined as

$$p(c, s|M_k) = \int_{\delta_k} p\left(c|s, \delta_k, M_k\right) p(s|\delta_k, M_k) p(\delta_k|M_k) d\delta_k; \quad k = 1, 2, \ldots 4. \tag{39}$$

In the expression for $p(c, s|M_k)$ in (39), δ_k denotes the vector of unobservables that characterize model M_k, $p\left(c|s, \delta_k, M_k\right)$ denotes the joint density for the option prices under M_k, $p(s|\delta_k, M_k)$ denotes the joint density for the spot prices under M_k and $p(\delta_k|M_k)$ denotes the prior density for δ_k under M_k. For models M_1, M_2 and M_3, δ_k includes the vector of variances. Hence, for these models, $p(\delta_k|M_k)$ is equal to the product of the density for the variances, given the parameters, and the prior density for the parameters.

In the present context, there is no closed form expression for $p(c, s|M_k)$. Various numerical approaches to the estimation of marginal likelihoods have been proposed; see Geweke (1999). However, such numerical procedures would be particularly burdensome in the present context, in particular due to the presence of the $n-$dimensional vector of latent variances, v, in both the reference model and two of the alternative models, M_2 and M_3. Fortunately, the Bayes Factors for M_2 and M_3, relative to Heston model, M_1, can be expressed in a particularly simple analytical form. This particular form of (38) is referred to as the Savage-Dickey (SD) density ratio; see Verdinelli and Wasserman (1995) and Koop and Potter (1999). For the case of M_2 versus M_1, the SD density ratio is given by:

$$SD_{21} = \frac{p(\lambda = 0|M_1, c, s)}{p(\lambda = 0|M_1)}, \tag{40}$$

where $p(\lambda = 0|M_1, c, s)$ denotes the ordinate of the marginal posterior for λ under the Heston model, M_1, evaluated at $\lambda = 0$ and $p(\lambda = 0|M_1)$ denotes the ordinate of the marginal prior for λ under M_1, evaluated at $\lambda = 0$. For (40) to be an equivalent representation of the Bayes Factor for M_2 versus M_1, BF_{21}, it must hold that

$$p(\delta_2|M_1, \lambda = 0) = p(\delta_2|M_2), \tag{41}$$

where $p(\delta_2|M_1, \lambda = 0)$ denotes the density for δ_2 in the model M_1 with $\lambda = 0$ imposed. The condition in (41) can be interpreted as the requirement that the prior distribution over all unknowns in the Heston model, M_1, conditional on $\lambda = 0$, is equivalent to the prior assigned to these unknowns in the submodel, M_2, in which $\lambda = 0$ is imposed from the outset. The assumption of *a priori* independence between λ and all other parameters in the Heston model (see (13)), means that this condition is satisfied.

For analogous reasons, the Bayes Factor for the case of M_3 ($\rho = 0$) versus M_1 reduces to

$$SD_{31} = \frac{p(\rho = 0|M_1, c, s)}{p(\rho = 0|M_1)}. \tag{42}$$

The denominators in both SD_{21} and SD_{31} are readily computed, given the specification of proper prior densities over both λ and ρ. The numerators are also easily computed via the MCMC simulation output, from which estimates of the marginal densities of λ and ρ are computed.

Adopting this same approach to compute the Bayes Factor for the BS model, M_4, versus M_1 is, however, problematic. This can be seen as follows. Under M_4, neither κ^a nor ρ is identified, with the model implying a constant variance of θ^a. Applying the results of Koop and Potter (1999), as long as

$$p(\theta^a, \mu|M_1, \sigma_v = 0) = p(\theta^a, \mu|M_4), \tag{43}$$

where θ^a and μ are the parameters common to both models, the Bayes Factor for M_4

versus M_1 still reduces to a form which is directly analogous to (40) and (42), namely,

$$SD_{41} = \frac{p(\sigma_v = 0|M_1, c, s)}{p(\sigma_v = 0|M_1)}. \tag{44}$$

Since (43) seems to be a reasonable assumption, it would appear, at first glance, that estimation of a Bayes Factor to test the BS model could proceed along the same lines as estimation of the Bayes Factors used to test M_2 and M_3. This is not the case however, for the following reason. Since σ_v must be positive, any prior density specified for σ_v must have an ordinate of zero when $\sigma_v = 0$, thereby producing a zero value for the denominator in (44). The same point applies to the posterior ordinate in the numerator. Hence, application of l'Hopital's rule would be required in order to evaluate SD_{41}. However, in practice, whilst the prior ordinate would be specified exactly as zero, the posterior ordinate would only ever be estimated numerically, as would the derivatives needed for the application of l'Hopital's rule. Hence, it is not possible to produce a reliable estimate of SD_{41}. We choose therefore not to calculate a Bayes Factor for M_4 versus M_1, ranking M_4 solely via its fit, predictive and hedging performance relative to the other models.

3.3.2 Fit and Predictive Performance

For model M_k with parameter vector δ_k, the residual associated with fitting the ith option price, C_i, is given by

$$res_i = C_i - [\beta_0 + \beta_1 q(z_i, \phi_k)], \quad i = 1, 2, \ldots, N, \tag{45}$$

where $q(z_i, \phi_k)$ denotes the theoretical option price associated with model M_k, $k = 1, 2, \ldots 4$, and $\delta_k = \{\phi_k, \mu, \beta, \sigma_u\}$. For M_1 the appropriate price is $q_H(z_i, \phi_1)$, as defined in (8). For M_2 and M_3, the price is $q_H(z_i, \phi_2)$ and $q_H(z_i, \phi_3)$ respectively; that is, the Heston option price, but with $\lambda = 0$ and $\rho = 0$ respectively imposed. For M_4, the price is the theoretical BS option price, defined as

$$q_{BS}(z_i, \phi_4 = \theta^a) = S_i \Phi(d_1) - K_i e^{-r_i \tau_i} \Phi(d_2), \tag{46}$$

where

$$d_1 = \frac{\ln(S_i/K_i) + (r_i + \theta^a/2)\tau_i}{\sqrt{\theta^a \tau_i}}, \tag{47}$$

$$d_2 = \frac{\ln(S_i/K_i) + (r_i - \theta^a/2)\tau_i}{\sqrt{\theta^a \tau_i}} \tag{48}$$

and $\Phi(x)$ denotes the cumulative normal distribution function evaluated at x; see, for example, Hull (2000). The quantity res_i is a nonlinear function of the underlying parameters and latent volatilities contained in ϕ_k, as well as a function of $\beta = (\beta_0, \beta_1)'$. Hence, its posterior distribution can be derived via the appropriate transformation of the posterior distribution of ϕ_k and β. As outlined in Section 3.2.6, $p(\beta|\sigma_u, \phi_k, c, s)$ is normal, with mean and variance given in (36) and (37) respectively. As such, the conditional posterior for res_i, $p(res_i|\sigma_u, \phi_k, c, s)$, is also normal, with mean

$$E(res_i|\sigma_u, \phi_k, c, s) = C_i - \left[\widehat{\beta}_0 + \widehat{\beta}_1 q(z_i, \phi_k)\right]$$

and variance

$$var(res_i|\sigma_u, \phi_k, c, s) = \sigma_u^2 \left[1, q(z_i, \phi_k)\right] (X'X)^{-1} \left[1, q(z_i, \phi_k)\right]',$$

with $\widehat{\beta}_0$ and $\widehat{\beta}_1$ the elements of the two-dimensional vector in (36) and X as defined following (37). As such, the marginal posterior of res_i is given by

$$p(res_i|c, s) = \int_{\sigma_u} \int_{\phi_k} p(res_i|\sigma_u, \phi_k, c, s) p(\sigma_u, \phi_k|c, s) d\sigma_u d\phi_k,$$

which can, in turn, be estimated via B MCMC draws of σ_u and ϕ_k, $\sigma_u^{(j)}$ and $\phi_k^{(j)}$, $j = 1, 2, \ldots, B$, as

$$p(res_i|c, s) = \frac{1}{B} \sum_{j=1}^{B} p(res_i|\sigma_u^{(j)}, \phi_k^{(j)}, c, s).$$

For M_1, the appropriate algorithm is the full MCMC scheme as described in Section 3.2. For M_2 and M_3, the algorithm is a reduced version of that scheme, with $\lambda = 0$ and $\rho = 0$ respectively imposed. For M_4, iterates of $\delta_4 = \{\theta^a, \mu, \beta, \sigma_u\}$ are generated from the posterior density for δ_4, given by

$$p(\delta_4|c, s) \propto p(c|s, \theta^a, \beta, \sigma_u) \times p(s|\theta^a, \mu) \times p(\theta^a, \mu),$$

where $p(c|s, \theta^a, \beta, \sigma_u)$ is given by

$$p(c|s, \theta^a, \beta, \sigma_u) = (2\pi)^{-N/2} \sigma_u^{-N} \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma_u^2}[C_i - [\beta_0 + \beta_1 q_{BS}(z_i, \theta^a)]]^2\right)$$

and $p(s|\theta^a \mu)$ is defined in accordance with

$$\begin{aligned} p(\ln s|\theta^a, \mu) &= (2\pi)^{-n/2} (\theta^a \Delta t)^{-n/2} \times \\ & \exp\left\{(-1/[2\theta^a \Delta t]) \sum_{t=2}^{n} (\ln S_t - [\ln S_{t-1} + (\mu - 0.5\theta^a)\Delta t])^2\right\}. \quad (49) \end{aligned}$$

The prior density, $p(\theta^a, \mu)$, is specified as being uniform over both parameters, truncated from below at $\theta^a = 0$. Draws of δ_4 can be produced via the conditionals for μ and θ^a respectively. Since the conditional for μ is normal, with mean and variance given respectively by

$$\widehat{\mu} = \frac{\sum_{t=2}^{n}(\ln S_t - \ln S_{t-1} + 0.5\theta^a \Delta t)}{n-1} \times \frac{1}{\Delta t}$$

and

$$\sigma_\mu^2 = \left(\frac{\theta^a}{n-1}\right) \times \frac{1}{\Delta t},$$

draws for μ given θ^a are readily obtained. Draws of θ^a conditional on μ can be obtained via a simple approximation of the one-dimensional conditional distribution function for θ (i.e. via 'Griddy Gibbs'), with the boundary at $\theta^a = 0$ imposed on the draws.

The proportion of 50% and 95% HPD intervals that cover zero can be calculated for each model M_k, with the best fitting model being the one for which this proportion is the closest to the nominal level. If the observed option prices used to define (45) belong to the vector c, the fit assessment is *within-sample.* If not, the fit assessment is *out-of-sample.* The latter is the form of fit assessment used in the empirical section.

Given the distributional assumption in (9), for model M_k the predictive density for an out-of-sample option price, C_f say, is given by:

$$p(C_f|c,s) = \int_{\delta_k} p(C_f|c,s,\phi_k,\beta,\sigma_u)p(\delta_k|c,s)d\delta_k, \tag{50}$$

where $p(C_f|c,s,\phi_k,\beta,\sigma_u)$ is a normal density with mean $\beta_0 + \beta_1 q(z_f, \phi_k)$ and variance σ_u^2 and $p(\delta_k|c,s)$ is the joint posterior density for parameters of model M_k. The notation z_f is used to denote the known factors associated with the future option contract f. In order to impose the lower bound that C_f must exceed in order for arbitrage opportunities to be avoided, in the construction of (50) we specify $p(C_f|c,s,\phi_k)$ as the *truncated* normal density

$$p(C_f|c,s,\phi_k,\beta,\sigma_u) = \frac{(2\pi\sigma_u^2)^{-1/2}}{(1-\Phi(slb_f))} \exp\left(-\frac{1}{2\sigma_u^2}[C_f - [\beta_0 + \beta_1 q(z_f,\phi_k)]]^2\right), \tag{51}$$

where

$$slb_f = \frac{lb_f - [\beta_o + \beta_1 q(z_f,\phi_k)]}{\sigma_u}$$

is the standardized version of the no-arbitrage lower bound,

$$lb_f = \max\{0, S_f - e^{-r_f \tau_f} K_f\}.$$

Again using the simulation output from the algorithm appropriate to model M_k, repeated draws from $p(\delta_k|c, s)$, $\delta^{(j)}$, $j = 1, 2, \ldots, B$, can be used to construct an estimate of $p(C_f|c, s)$ as

$$\widehat{p(C_f|c, s)} = \frac{1}{B}\sum_{j=1}^{B} p(C_f|c,s, \phi_k^{(j)}, \beta^{(j)}, \sigma_u^{(j)}). \tag{52}$$

In Section 4, prediction intervals constructed from (52) are used to rank the predictive performance of the models.

3.3.3 Hedging Performance

An important measure of the performance of the alternative volatility models is the extent to which they produce small hedging errors. In this paper we focus on the errors associated with single instrument hedge portfolios, in which movements in the underlying spot price, S_t, are hedged against by taking the appropriate position in a single option contract, with price C_t, at time t. In this case the resulting cash position with minimum variance is

$$C_t - N_k S_t, \tag{53}$$

where N_k denotes the number of shares in the underlying asset in which the investor goes long for every call option in which the investor goes short, assuming model M_k. In the case of the Heston model, M_1, N_1 is defined as

$$N_1 = \frac{\partial q_H}{\partial S_t} + \frac{\rho \sigma_v}{S_t}\frac{\partial q_H}{\partial v_t},$$

where

$$\frac{\partial q_H}{\partial S_t} = P_1$$

and

$$\frac{\partial q_H}{\partial v_t} = S_t \frac{\partial P_1}{\partial v_t} - K e^{-r_t \tau} \frac{\partial P_2}{\partial v_t}, \tag{54}$$

with q_H as defined in (8) and P_1 and P_2 as defined in the appendix; see Bakshi, Cao and Chen (1997) and Chernov and Ghysels (2000). The hedging error over one day, say, for the Heston model, H_1, is defined as the difference between the minimum variance hedge portfolio, as constructed on day t and invested at the risk-free rate r_t, and the value of that same portfolio when unwound one day later, namely,

$$H_1 = (C_t - N_1 S_t)\, e^{r_t \tau} - (C_{t+1} - N_1 S_{t+1})\,, \tag{55}$$

where C_{t+1} and S_{t+1} denote respectively the option and spot prices on day $t+1$. Since H_1 is a function of the unknown parameters and variances, via N_1, the posterior density of H_1, $p(H_1|c,s)$, can be estimated from the MCMC output associated with estimation of the Heston model. That is, draws of ϕ_1 can be used to produce draws of H_1, which can then be used to produce a nonparametric kernel estimate of $p(H_1|c,s)$. The distribution of hedging errors over any time period can be constructed in an analogous way.

The same sort of exercise can be performed for each of the alternative models, M_2, M_3 and M_4. For the first two of these alternative models, the relevant posterior densities for the hedging errors, $p(H_2|c,s)$ and $p(H_3|c,s)$, can be estimated from the output of the MCMC schemes in which $\lambda = 0$ and $\rho = 0$ are imposed respectively. In the case of M_4, specification of

$$N_4 = \Phi\,(d_1)\,,$$

with d_1 as defined in (47), renders the portfolio in (53) delta-hedged; see Hull (2000). For the three stochastic volatility models, the derivatives with respect to v_t which enter (54) need to be computed numerically. The best performing model according to this criterion is the model with the hedging error density most closely concentrated around zero.

3.3.4 Model Averaging

As described in Section 3.3.1, the posterior probability of three of the four alternative models can be estimated from the option prices. These posterior probabilities can be used to produce a model-averaged predictive density, which can, in turn, be used as the tool for prediction rather than the predictive associated with any one particular model.[16]The rationale of this approach is that with option prices being determined by the interaction of market participants using different distributional assumptions, the model-averaged predictive may well have better coverage properties that the predictives associated with specific models. Given model-specific predictive densities, $p(C_f|M_k,c,s)$, $k = 1,2,3$, the averaged predictive density, $p_a(C_f|c,s)$, is defined as:

$$p_a(C_f|c,s) = \sum_{k=1}^{3} p(C_f|M_k,c,s)P(M_k|c,s), \tag{56}$$

with $P(M_k|c,s)$, $k = 1,2,3$ calculated as described in Section 3.3.1.[17]

4 Numerical Illustration: News Corporation Option Prices

4.1 Data Description

The methodology is demonstrated using data on News Corporation option and equity trades over the four year period: 1998 to 2001. All data has been obtained from the Australian Stock Exchange. Only options with maturities of between 15 and 90 days are included in the dataset. In order to use the option data to produce inferences on the day to day movements in the underlying spot prices process, options trades are selected from one particular period of time each day, namely the last hour of trading. A range of option prices that maximizes the moneyness spread is then selected from the set of prices observed during this period. A total of 10904 observations are used for estimation, with a maximum of 60 prices selected from any one day in the within-sample period. The out-of-sample period constitutes the nine trading days from 11 December, 2001 to 21 December 2001, with 855 trades used for the out-of-sample assessments. Since equity trades on News Corporation stock occur very frequently, for each option trade it is possible to obtain a virtually simultaneous equity price: usually recorded within a few seconds of the option trade. When several equity trades are recorded at exactly the same time, a weighted average is taken, with the weights determined by the trading volume. Observations on the discretized process for equity prices are taken as the average of the spot prices observed during the last hour of normal trading. The dividends paid on News Corporation shares are typically about 0.1% of share value, paid 6-monthly. The impact of dividends on share prices is therefore so small that they have only been taken into account as a constant continuous discount factor.

4.2 Numerical Results

4.2.1 Posterior Results for the Heston Model

Table 1 presents point and interval estimates of both the parameters and selected variances associated with the Heston model. All results are produced by running the MCMC algorithm for 5000 iterations, with the first 500 iterations discarded. The starting values for the parameters are as follows: $\kappa^a = 2.0$; $\theta^a = 0.20$; $\sigma_v = 0.55$; $\lambda = 0.5$; $\rho = -0.05$ and $\mu = 0.16$.[18]The prior mean and standard deviation for λ are respectively 0.2 and 1.0. The prior mean and standard deviation for ρ are respectively -0.2 and 0.3.[19]The prior specifications and starting values for the parameters are determined by a combination of preliminary analysis of New Corporation returns data and preliminary experimentation with the MCMC algorithm. In order to improve the acceptance rate of the component of the MCMC algorithm related to v, the elements of the full vector v are selected in blocks with an average size of 10. The actual block lengths at each iteration are chosen randomly; see Shepherd and Pitt (1997) and Strickland, Forbes and Martin (2003) for details.

The acceptance rates for the four MH subchains vary considerably. The subchains for the parameters λ and ρ have acceptance rates of 100% and 91% respectively. The acceptance rates for κ^a, θ^a and σ_v are respectively 64%, 21% and 48%, and that of the (blocked) vector v, 14%. The normal candidate densities for both λ and ρ, constructed as described in Sections 3.2.3 and 3.2.4, are updated only when v changes, since it has been determined in preliminary experimentation that v is the most important determinant of the form of the conditionals for these two parameters. This represents a considerable computational saving.

Table 1 reports the mean, mode and (approximate) 95% Highest Posterior Density (*HPD)* estimates for the parameters and selected random variances[20]. The point estimates of κ^a correspond to daily persistence measures of $(1 - 3.240/252) = 0.987$ and $(1 - 3.483/252) = 0.986$ respectively, with the interval estimate translating into a (daily) persistence interval of (0.983, 0.988). All estimates are thus well within the realm of typical returns-based estimates of such measures. For instance, the persistence measure from a Generalized Autoregressive Conditional Heteroscedastic (GARCH) (1,1) model estimated for News Corporation returns over the 1998-2001 period is 0.968. The point estimates of θ^a correspond to an estimate of long-run volatility of approximately $\sqrt{0.210} = 0.458$, a value that corresponds very closely to the unconditional mean of volatility associated with a GARCH(1,1) model estimated over the 1998-2001 period, namely 0.457. The point and interval estimates of λ all indicate a positive value, which contrasts with the negative values that are typical found for this parameter. That said, empirical estimates of λ are notoriously variable across different samples of option prices, with positive estimates having been reported; see, for example, Guo (1998). The estimates of σ_v imply a moderate degree of excess kurtosis in the returns process, whilst the estimates of ρ indicate that negligible returns skewness is implied by the joint options and spot datasets. The point estimates of μ imply an (annualized) rate of return on the News Corporation stock of approximately 30%, well in excess of the returns-based mean rate of return of 15% estimated for the 1998-2001 period. However, the interval estimate indicates the extent of the uncertainly associated with estimation of this parameter, with 95% probability assigned to the range $(-0.170, 0.730)$.

The point estimates of the variances reported in Table 1 are slightly lower than the estimated value of the long-run mean, θ^a. However, the interval estimates cover values that overlap with the interval estimates of θ^a.

Figure 1 graphs the marginal densities for the parameters and Figure 2 the marginal posteriors for selected random variances.

Table 1 here.

Figure 1 here.

Figure 2 here.

4.2.2 Model Ranking

In this section we apply the methods discussed earlier to rank the three restricted models, M_2, M_3 and M_4, and the full model M_1. From the interval estimates reported in Table 1, it is evident that two of the three restricted models, M_2 and M_4, are clearly rejected by the data, with neither of the intervals covering the relevant parameter restrictions. The interval estimate for ρ also fails to cover the value of $\rho = 0$, thereby providing evidence against M_3. However, as is evident from the marginal posterior for ρ in Figure 1, there is substantial posterior mass in the region very close to zero.

To supplement these results, we estimate the three submodels, assessing the out-of-sample performance of each, relative to that of M_1, along the lines discussed in Sections 3.3.2 and 3.3.3, as well as constructing Bayes Factors for M_2 and M_3. The details of the algorithm used to estimate M_2 and M_3 are identical to the details given in the Section 3.2, apart from the obvious parameter restrictions associated with the nested models. Estimation of the BS model, M_1, in the manner described in Section 3.3.2, produces a marginal posterior for $\sqrt{\theta^a}$ that is extremely concentrated around a single value, namely $\sqrt{\theta^a} = 0.37$. With only this parameter affecting the theoretical option prices, we report the fit, predictive and hedging results conditional on this single value. As noted in Section 3.3.1, we choose not to construct a Bayes Factor for this submodel.

Table 2 reports the estimated Bayes Factors for M_2 and M_3, with M_1 as the reference model. The bottom row in the table gives the corresponding model probabilities, based on equal prior probabilities for all three models and assuming that these three models span the model set. It is clear that the posterior probabilities support M_1, but with substantial weight also assigned to M_3. This latter result tallies with the substantial posterior weight assigned to the region around $\rho = 0$ by the marginal posterior of ρ derived from the full model, M_1. Negligible posterior weight is assigned to M_2, again a result which tallies with the clear non-zero estimates of λ in the full model.

Table 2 here.

The fit and predictive results reported in Table 3 represent the proportion of times that each criterion is satisfied, for each model, for the 855 out-of-sample observations. If the model is correctly specified, this proportion should approximate the nominal coverage level. As is evident, models M_1, M_2 and M_3 are all very close to the nominal level in the case of the 95% fit intervals and slightly overstate the nominal level in the case of the 95% prediction intervals. All three of these models understate the 50% nominal coverage of both the fit and prediction intervals. The coverage of the prediction intervals for all four models is always greater than that of the corresponding fit intervals, since the prediction intervals factor in the extra variation associated with the future option price, as captured by σ_u^2.

The fit and predictive results are consistent with the posterior model probabilities to the extent that M_1 out performs M_3, but not by a substantial amount. Interestingly, despite the negligible probability weight assigned to M_2, this model does best (in terms of closeness to the nominal coverage level) for three of the four criterion reported below. The Black-Scholes model, M_4, is clearly inferior to all other models in terms of all four criteria, with none of the fit intervals, at either nominal coverage level, encompassing the observed out-of-sample option prices. The coverage of the predictions intervals is also markedly below the nominal level, indicating that the model is significantly misspecified.

The model-averaged predictive, constructed according to (56), with $P(M_k|c,s)$, $k = 1,2,3$, as given in Table 2, has an interquartile coverage of 0.316 and a 95% interval coverage of 0.986. Since only M_1 and M_3 have non-negligible posterior probability, the weighting effectively occurs with respect to the predictives of these models only. For the 95% interval in particular, there is negligible difference between the coverage of the model-averaged predictive and that of the individual models. For the interquartile interval, the averaging produces marginally better coverage than that of both individual intervals.

Table 3 here.

In Table 4 the hedging error results associated with the four models are reported. Hedging errors are calculated using (55) for M_1 and the version of (55) appropriate for the remaining models, as described in the text immediately following (55). Errors are calculated for one day and five days ahead, with the portfolio constructed at the end of the estimation period and not rebalanced during the entire out-of-sample period. On each of these days the hedging errors associated with all contracts are calculated. The errors are then averaged across all contracts.[21]It is these averaged hedging errors to which the summary statistics in Table 4 relate and whose posterior densities are graphed in Figure 3.

Table 4 here.

The results in Table 4 make it clear that there is very little difference between all four models according to this criterion. It is also clear that none of the *HPD* intervals cover zero. That said, for all four models, the errors associated with the hedge portfolio one day out from its construction are minimal, in terms of both the point and interval estimates. These errors represent between approximately 1% of the average magnitude of the option prices on the first out-of-sample day. As would be anticipated, the errors increase over time without any rebalancing taking place, with the five-day out hedging errors representing approximately 5% of the average magnitude of the option prices on the fifth out-of-sample day.

Figure 3 here.

In Figure 3 the posterior densities for the one day ahead and five days ahead hedging errors associated with model M_1, M_2 and M_3 are presented. Densities are not constructed for M_4 since the hedging errors are calculated only for a single value of the volatility parameter.

5 Conclusions

In this paper a new methodology for producing option and spot price-based estimates of the parameters of a stochastic volatility model is presented. The method has been developed within the context of the Heston (1993) theoretical option pricing model and certain variants thereof. The numerical scheme adopted exploits the state-space representation of the Heston spot and volatility process, as well as the approximate linearity of the relationship between the theoretical option price and the price of volatility risk. Simulation of the latent volatilities occurs via the application of a bivariate Kalman filter and smoother, with information from the option prices impacting on the filter via a measurement equation in which the BS implied volatilities proxy the option prices. Construction of Bayes Factors, in addition to fit, prediction and hedging intervals, enables the alternative variants of the proposed model to be ranked on the basis of observed option and spot price data.

Application of the methodology to Australian News Corporation stock and options data produces estimates of the parameters of the Heston model that imply a very persistent volatility process, with the degree of volatility in volatility indicating that a certain amount of excess kurtosis characterizes the spot price data and/or has been factored into the options data. Skewness is not a feature of the results, with the posterior probability associated with the volatility model in which zero skewness is imposed being only slightly less than that associated with the unrestricted Heston model. The model that imposes a zero premium for volatility risk is clearly rejected, as it is assigned virtually zero posterior probability. Both point and interval estimates of the relevant parameter indicate that the risk premium that is factored into option prices over this period is positive. Hence, the risk-neutral volatility process is estimated to converge more rapidly to a lower long-run mean than would be the case for the objective process. This implies, in turn, that the observed option prices are lower than would be the case had they been priced under the objective measure. The constant volatility BS model is also clearly rejected by the data in the sense that the point and interval estimates of the variance of the Heston volatility process are non-zero. The within-sample fit and out-of-sample predictive performance of the BS model is also markedly inferior to that of all variants of the Heston model that are considered. On the other hand, there is little to choose between the stochastic volatility variants in terms of fit, predictive and hedging performance. Given the similarity in the predictive results across models, the model averaging process produces only a minor improvement in predictive performance.

References

[1] Bakshi, G., C. Cao and Z. Chen. (1997). "Empirical Performance of Alternative Option Pricing Models." Journal of Finance 52, 2003-2049.

[2] Bates, D.S. (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in PHLX Deutsche Mark Options." The Review of Financial Studies 9, 69-107.

[3] Bates, D.S. (2000). "Post-87 Crash Fears in the S&P 500 Futures Option Market." Journal of Econometrics 94, 181-238.

[4] Black, F. and M. Scholes. (1973). "The Pricing of Options and Corporate Liabilities." Journal of Political Economy 81, 637-659.

[5] Chernov, M. and E. Ghysels. (2000). "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation." Journal of Financial Economics 56, 407-458.

[6] Chib, S. and E. Greenberg. (1996). "Markov chain Monte Carlo simulation methods in Econometrics." Econometric Theory 12, 409-431.

[7] Cox, J.C., J.E. Ingersoll and S.A. Ross. (1985). "A Theory of the Term Structure of Interest Rates." Econometrica *53, 385-408.*

[8] De Jong, P. and N.G. Shephard. (1995). "The Simulation Smoother for Time Series Models." Biometrika 82, 339-350.

[9] Eraker, B. (2003). "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices." Draft Paper, Duke University. Forthcoming Journal of Finance.

[10] Eraker, B., M.J. Johannes N.G. and Polson. (2003). "The Impact of Jumps in Returns and Volatility." Journal of Finance 53, 1269-1300.

[11] Geweke, J. (1999). "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication." Econometric Reviews 18, 1-73.

[12] Guo, D. (1998). "The Risk Premium of Volatility Implicit in Currency Options." Journal of Business and Economic Statistics 16, 498-507.

[13] Heston, S.L. (1993). "A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." The Review of Financial Studies 6, 327-343.

[14] Hull, J.C. (2000). Options, Futures, and Other Derivative Securities. 4rd ed., New Jersey: Prentice Hall.

[15] Hull, J.C. and A. White. (1987). "The Pricing of Options on Assets with Stochastic Volatilities." Journal of Finance 42, 281-300.

[16] Jones, C. (2003). "The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets." Journal of Econometrics 116, 181-224.

[17] Koop, G. and S. Potter. (1999). "Bayes Factors and Nonlinearity: Evidence from Economic Time Series." Journal of Econometrics 88, 251-281.

[18] Pan, J. (2002). "The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-series Study." Journal Of Financial Economics 63, 3-50.

[19] Shephard, N. and M. Pitt. (1997). "Likelihood Analysis of Non-Gaussian Measurement Time Series." Biometrica 84, 653-667.

[20] Strickland, C.M., C.S. Forbes and G.M. Martin. (2003). "Bayesian Analysis of the Stochastic Conditional Duration Model." Working Paper 14/2003, Monash University.

[21] Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions." The Annals of Statistics 22, 1701-1762.

[22] Verdinelli, I. and L. Wasserman. (1995). "Computing Bayes Factors Using a Generalization of the Savage-Dickey Ratio." Journal of the American Statistical Association 90, 614-618.

Appendix: Solution of the Heston model

The Heston stochastic volatility model is based on the bivariate stochastic process in (3) and (4). From this process, the partial differential equation for the option price, U say,

$$\begin{aligned} & \frac{1}{2}v(t)S^2(t)\frac{\partial^2 U}{\partial S(t)^2} + \rho\sigma_v v(t)S(t)\frac{\partial^2 U}{\partial S(t)\,\partial v(t)} + \frac{1}{2}\sigma_v^2 v(t)\frac{\partial^2 U}{\partial v(t)^2} \\ & + rS(t)\frac{\partial U}{\partial S(t)} + [\kappa^a(\theta^a - v(t)) - \lambda v(t)]\frac{\partial U}{\partial v(t)} - rU + \frac{\partial U}{\partial t} \\ = & \ 0 \end{aligned} \qquad (57)$$

can be obtained largely by the standard Black-Scholes type procedure. Ito's lemma is used to obtain an expression for the change in the value of the option, dU, and in the underlying, $dS(t)$, and an appropriate combination formed to attempt to eliminate randomness. Since there are two sources of randomness and only one hedging instrument, not all randomness can be removed. If we wish to remove all terms involving $d\varepsilon_1(t)$ and leave the term in $d\varepsilon_2(t)$ untouched, we form the combination

$$dU - \frac{\partial U}{\partial S(t)}dS(t). \qquad (58)$$

In the absence of other random terms, this could be equated with growth at the risk-free interest rate r,

$$r(U - \frac{\partial U}{\partial S(t)}S(t))dt. \qquad (59)$$

However, since there is no tradable security relating to the stochastic volatility, the expression for (58) obtained using the Ito lemma still includes a random term with expected value zero, which cannot be hedged away, namely $\sigma_v\sqrt{v(t)}\frac{\partial U}{\partial v(t)}d\varepsilon_2(t)$. Following Cox, Ingersoll and Ross (1985), this random term can be replaced by a term $-\lambda v(t)\frac{\partial U}{\partial v(t)}dt$, yielding (57). Here, $\lambda v(t)\frac{\partial U}{\partial v(t)}$ is the premium associated with the volatility risk, and we call λ the risk premium parameter. Note that if the investor were neutral with respect to volatility risk, and therefore did not require a premium, λ would be zero. In equilibrium, the risk premium term $\lambda v(t)\frac{\partial U}{\partial v(t)}$ is equal to the excess expected return over the risk-free rate demanded by the investor as a result of volatility risk.

An alternative approach is to decompose $d\varepsilon_2(t)$ into a $d\varepsilon_1(t)$ component, and a component independent of $d\varepsilon_1(t)$,

$$d\varepsilon_2(t) = \rho d\varepsilon_1(t) + \sqrt{1-\rho^2}dW, \tag{60}$$

where dW is a Gaussian process independent of $d\varepsilon_1(t)$; see, for example, Chernov and Ghysels (2000). Then a different partially hedged portfolio is formed where $d\varepsilon_1(t)$ components from both sources are eliminated. The appropriate portfolio is the one given in (53). In this case, to obtain the Heston equation we must replace the term

$$\sigma_v \frac{\partial U}{\partial v(t)} \left[\rho(r-\mu)dt + \sqrt{1-\rho^2}\sqrt{v(t)}dW \right] \tag{61}$$

by $-\lambda v(t)\frac{\partial U}{\partial v(t)}dt$.

Equation (57) can be rewritten in the form

$$\begin{aligned} & \frac{1}{2}v(t)S^2(t)\frac{\partial^2 U}{\partial S(t)^2} + \rho\sigma_v v(t)S(t)\frac{\partial^2 U}{\partial S(t)\,\partial v(t)} + \frac{1}{2}\sigma_v^2 v(t)\frac{\partial^2 U}{\partial v(t)^2} \\ & +rS(t)\frac{\partial U}{\partial S(t)} + \kappa(\theta - v(t))\frac{\partial U}{\partial v(t)} - rU + \frac{\partial U}{\partial t} \\ = & \; 0 \end{aligned} \tag{62}$$

where

$$\kappa = \kappa^a + \lambda, \;\; \theta = \frac{\kappa^a \theta^a}{\kappa^a + \lambda}. \tag{63}$$

This is the form of (57) which would arise in a hypothetical risk neutral world characterized by mean reversion parameter κ and long-run mean parameter θ. For positive values of λ, the variances in the risk neutral world would revert more rapidly to a lower mean.

The Heston option price, denoted by q_H in the text, is the solution of (57) subject to the boundary conditions

$$\begin{aligned} U(S(t), v(t), t) &= Max(0, S(t) - K) \\ U(0, v(t), t) &= 0 \end{aligned}$$

$$
\begin{aligned}
\frac{\partial U}{\partial S(t)}(\infty, v(t), t) &= 1 \\
U(S(t), \infty, t) &= 0
\end{aligned}
$$

and

$$
rS(t)\frac{\partial U}{\partial S(t)}(S(t), 0, t) + \kappa^a\theta^a\frac{\partial U}{\partial v(t)}(S(t), 0, t) - rU(S(t), 0, t) + U(S(t), 0, t) = 0.
$$

These conditions correspond to a European call option with strike price K and maturing at time T. For the detailed formula for q_H we refer to Heston (1993). Here we simply make a few brief comments on the solution. We denote by p^* the probability density function corresponding to the processes (1) and (2), which characterize the hypothetical risk neutral world. Transforming to $x(t) = \ln S(t)$, these equations become

$$
dx(t) = [r - \frac{1}{2}v(t)]dt + \sqrt{v(t)}d\varepsilon_1(t) \tag{64}
$$

and

$$
dv(t) = \kappa[\theta - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t) \tag{65}
$$

respectively. The risk neutral method of solution outlined above yields

$$
\begin{aligned}
q_H &= e^{-r\tau}\int_K^\infty (S(t) - K)\, p^*(S(t))\, dS(t) \\
&= S(t)P_1 - Ke^{-r\tau}P_2, \qquad (66)
\end{aligned}
$$

where

$$
P_1 = \int_K^\infty \frac{e^{-r\tau}S(t)}{S(t)}\, p^*(S(t))\, dS(t) \tag{67}
$$

and

$$
P_2 = \int_K^\infty p^*(S(t))\, dS(t). \tag{68}
$$

The quantity P_2 is the risk neutral probability that the option will be exercised. Since $S(t) = e^{-r\tau}E^*(S(t))$,

$$
\begin{aligned}
P_1 &= \int_K^\infty \frac{S(t)}{E^*(S(t))}p^*(S(t))dS(t) \\
&= \int_K^\infty f(S(t))\, dS(t)
\end{aligned}
$$

is also a probability. As shown in Heston (1993), it is the probability that the option will be exercised in a risk neutral world characterized by the processes

$$dx(t) = [r + \frac{1}{2}v(t)]dt + \sqrt{v(t)}d\varepsilon_1(t) \quad (69)$$

$$dv(t) = [\kappa\theta - (\kappa - \rho\sigma_v)v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t). \quad (70)$$

Although the above discussion indicates that the pdf p^* corresponding to the bivariate process in (64) and (65) is the appropriate risk neutral measure for the Heston model, it is not clear whether the transformation to the risk neutral measure is well-defined (Chernov and Ghysels, 2000). For very small values of the volatility, the risk premium on asset risk is very large, leading to arbitrage opportunities. Extremely small values of the volatility are unlikely in practice. Our numerical results, for instance, produce a mean value for the stochastic variance of approximately 0.2, a value that is fairly typical of the variance in certain empirical stock market data.

Acknowledgements

This research has been supported by a Large Australian Research Council Grant and a Monash University Research Grant. The authors would like to thank Vance Martin for some very helpful suggestions and comments on an earlier version of the paper.

Notes

1. More details of the derivation of q_H and of the interpretation of P_1 and P_2 are given in the Appendix.

2. On the basis of certain assumptions, Bates is able to estimate a lower bound for λ.

3. Jones (2003) also uses Bayesian methods to conduct option-based inference about a stochastic volatility model. However, his methodology is somewhat different from that proposed here, in that observed market option prices are not used in the inferential procedure.

4. Although Eraker exploits information in both the spot and option prices in the estimation procedure, the precise manner in which the spot and option prices processes feed into the 'single-move' MCMC algorithm that he proposes is not made explicit.

5. Results on the empirical usefulness of allowing for random jumps in the asset process and/or the volatility process are rather mixed; see, for example, Eraker (2003).

6. The assumption of a constant variance for the pricing errors is maintained, since, for the data set under consideration in the paper, no discernable pattern in the variance, across moneyness in particular, is found.

7. In the dataset used in the empirical analysis prices from the spot market are matched as closely as possible with the prices observed in the option market. Since the market for News Corporation shares is very liquid, this matching process is very accurate, with a spot price recorded usually within a few seconds of the option trade.

8. To rule out arbitrage, the distribution of C_i should strictly speaking be truncated at a lower bound of $lb_i = \max\{0, S_i - e^{-r_i \tau_i} K_i\}$; see Hull (2000). Since experimentation has established that the truncation has only minimal impact on inferences, we choose to ignore it in the estimation procedure. We do however, invoke the truncation when producing the predictive densities used in ranking alternative models.

9. See Bates (2000) for more on this issue.

10. These assumptions could be questioned, in particular given prior knowledge of the possible relationship between the signs of λ and ρ. They are however maintained for the sake of computational convenience.

11. Depending on the size of the sample to which the method is applied, the vector v may be further divided into smaller blocks, in order to increase the acceptance rates associated with this component of the algorithm. However, we describe the simulation of v below in terms of the full vector, with details of the further blocking that is applied in the context of the empirical application provided in Section 3

12. See Eraker, Johannes and Polson (2003) for evidence that the daily interval is small enough to render the discretization bias arising from (15) negligible.

13. The Kalman filter need not be rerun, only the smoothing algorithm.

14. These restrictions serve to impose mean reversion in the volatility process *a priori* and to reflect the fact that θ^a is the long-run mean of a variance process.

15. Note that a proper prior on λ is required for the purpose of constructing well-defined Bayes Factors related to λ.

16. See Geweke (1999) for discussion of the principles of Bayesian model averaging.

17. Model-averaged fit and hedging error densities can also be produced in a similar manner. We focus only on the model-averaged predictive density as it has a clear interpretation as an inferential tool and, hence, better serves to illustrate the potential benefits of model averaging.

18. As β and σ_u are essentially nuisance parameters, we do not devote space to them in the reporting of results in the text. However, we note that the posterior mean estimates of these parameters are respectively $(-0.075, 1.002)$ and 0.062.

19. These prior parameters imply only a very small prior probability of ρ falling beyond the bounds of ± 1. These bounds are imposed on the posterior distribution by way of discarding any draws of ρ that fall beyond them.

20. An *HPD* interval is one with the specified probability coverage, whose inner ordinates are not exceeded by any density ordinates outside the interval. The intervals reported in Table 1 and elsewhere in the paper are approximate *HPD* intervals in that the kernel smoothing procedure used to estimate the marginal posteriors does not always enable the ordinate condition described here to be satisfied exactly. Furthermore, for densities that are multimodal the 95% intervals are contructed so that the ordinates of the lower and upper bounds are as close as possible to being equal, subject to the restriction that the tail probabilities sum to 5%. This implies that there may be ordinates within the interval that are smaller than ordinates beyond the interval.

21. The data was not plentiful enough to construct meaningful hedging statistics for the different moneyness categories.

Table 1: Marginal Posterior Density Results for the Heston Model

	Mode	Mean	95% *HPD* interval
Parameter			
κ^a	3.240	3.483	(2.920, 4.300)
θ^a	0.213	0.221	(0.174, 0.258)
σ_v	0.640	0.651	(0.600, 0.710)
λ	1.900	1.539	(0.700, 2.100)
ρ	0.070	0.062	(0.010, 0.110)
μ	0.320	0.284	(-0.170, 0.730)
Variance			
v_{100}	0.159	0.164	(0.137, 0.193)
v_{500}	0.200	0.189	(0.163, 0.208)
v_{990}	0.163	0.163	(0.144, 0.183)

Table 2: Bayes Factors and Model Probabilities
Entry (i, j) indicates the Bayes Factor
in favour of M_j versus M_i

	M_1	M_2	M_3
M_1	1.000	2.342×10^{-21}	0.874
M_2		1.000	3.732×10^{20}
M_3			1.000
$P(M_k\|c, s)$	0.534	1.250×10^{-21}	0.466

Table 3: Fit and Predictive Performance Measures

Criterion[a]	M_1	M_2	M_3	M_4
Zero in Interquartile Fit Interval[b]	0.2012	0.2070	0.1485	0.000
Zero in 95% Residual Interval[b]	0.9520	0.9404	0.9450	0.000
C_f in Interquartile Predictive Interval[b]	0.3135	0.4339	0.2924	0.0491
C_f in 95% Predictive Interval[b]	0.9860	0.9848	0.9825	0.6339

(a) All figures represent proportions of 900.

(b) The $(1-\alpha)$% Interval is the interval which excludes $\alpha/2$% in the lower and upper tails of the fit/predictive distribution. This interval equals the $(1-\alpha)$% *HPD* interval only for those distributions which are symmetric around a single mode.

Table 4: Hedging Performance of the Different Models
Means of (Average) Hedging Error Densities with 95% *HPD* Intervals in Brackets ($)

M_1	M_2	M_3	$M_4^{(a)}$
	One Day Ahead		
-0.011	-0.011	-0.010	-0.006
(-0.012, -0.010)	(-0.012, -0.010)	(-0.0105, -0.00995)	n.a.
	Five Days Ahead		
0.063	0.063	0.065	0.062
(0.061, 0.064)	(0.061, 0.065)	(0.0645, 0.0652)	n.a.

(a) Since only a single point estimate of the BS volatility has been produced, *HPD* intervals cannot be constructed.

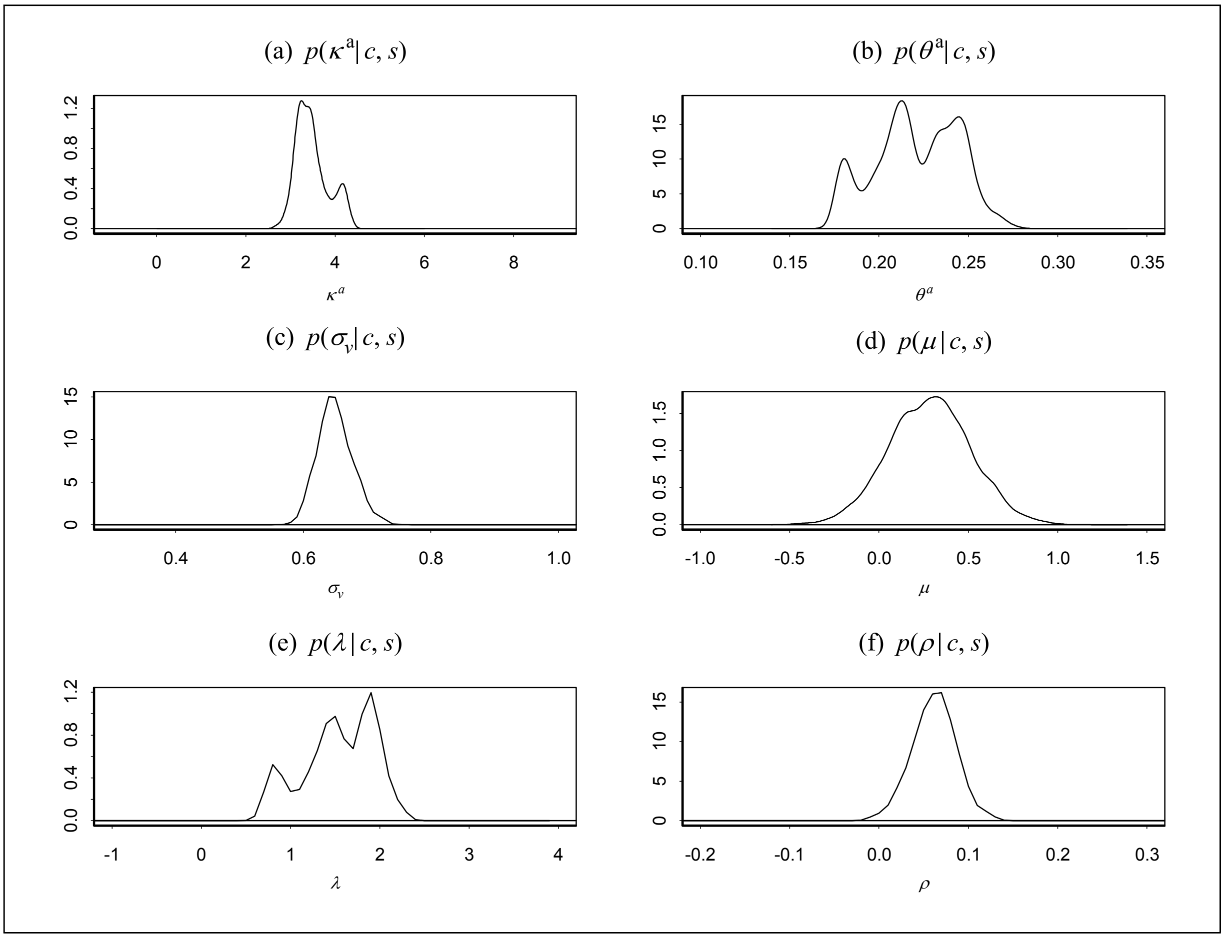

Figure 1: Marginal Posterior Densities for the Parameters of the Heston Model.

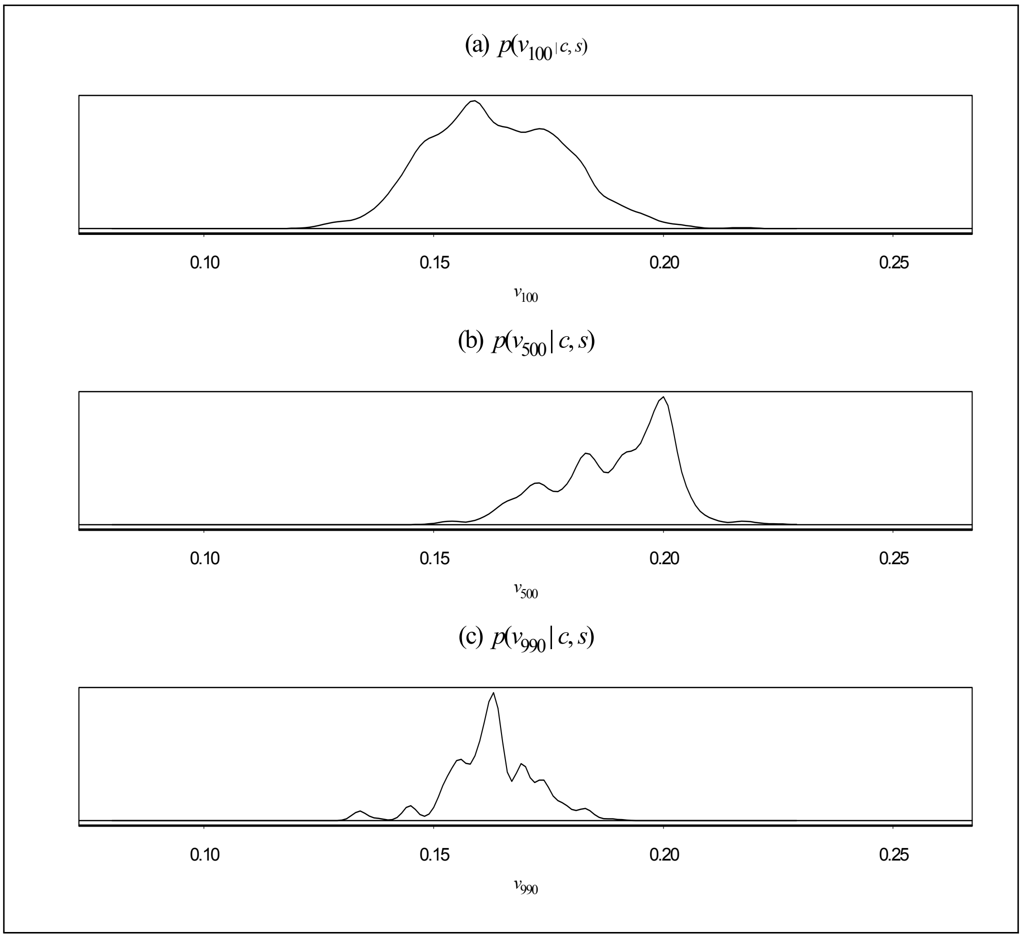

Figure 2: Marginal Posterior Densities for Selected Variances from the Heston Model.

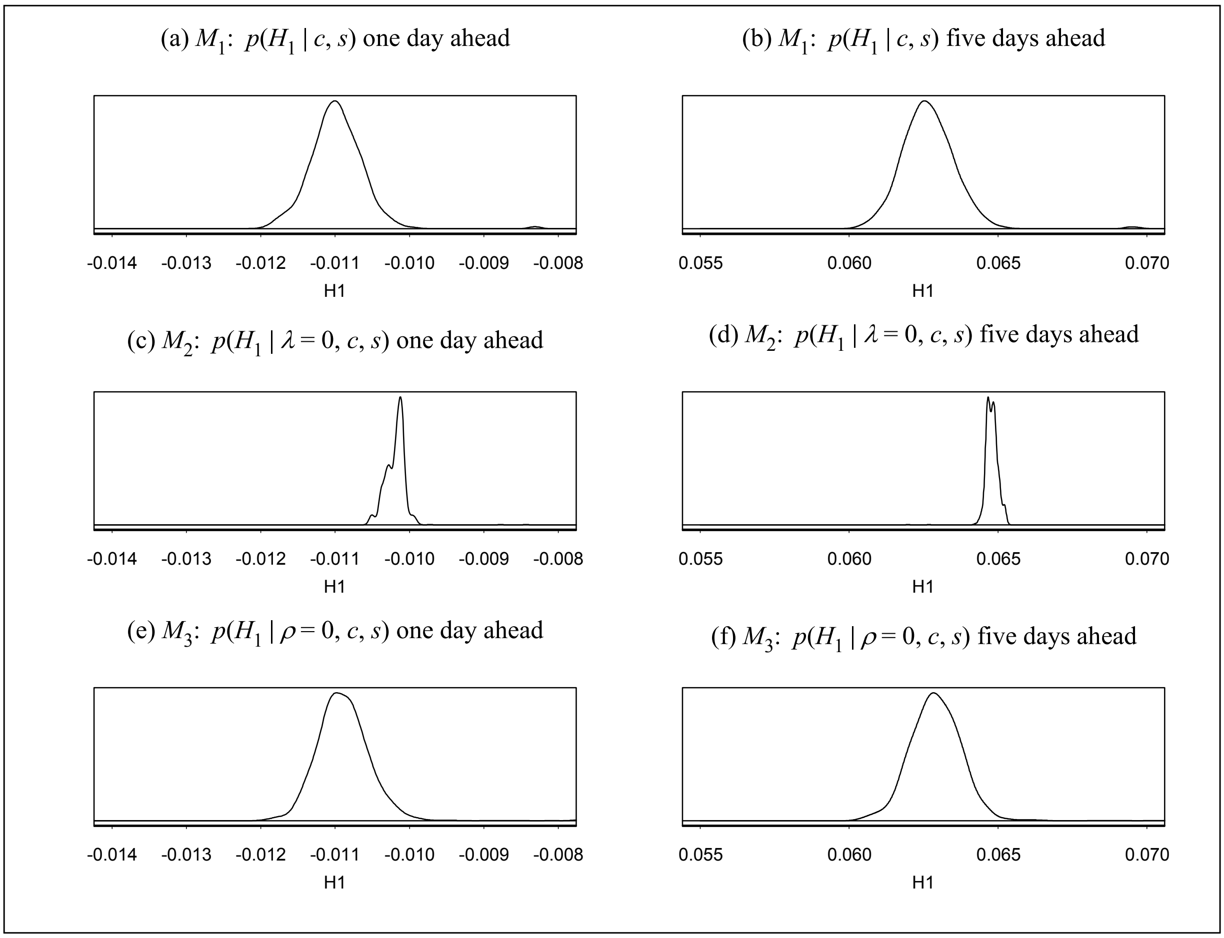

Figure 3: Hedging Error Densities for M_1, M_2 and M_3 : \$ value one day ahead and five days ahead.