ISSN 1440-771X

MONASH UNIVERSITY
AUSTRALIA

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

Bayesian Estimation of a Stochastic Volatility Model Using Option and Spot Prices

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

Working Paper 2/2002

Bayesian Estimation of a Stochastic Volatility Model Using Option and Spot Prices*

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

Department of Econometrics and Business Statistics, PO Box 11E

Monash University, Vic., 3800, Australia; Email: Gael.Martin@BusEco.monash.edu.au.

March, 2002

Abstract

In this paper we apply Bayesian methods to estimate a stochastic volatility model using both the prices of the asset and the prices of options written on the asset. Implicit posterior densities for the parameters of the volatility model, for the latent volatilities and for the market price of volatility risk are produced. The method involves augmenting the data generating process associated with a panel of option prices with the probability density function describing the dynamics of the underlying bivariate spot price and volatility process. Posterior results are produced via a hybrid Markov Chain Monte Carlo sampling algorithm. Candidate draws which assume a given dynamic process for the volatility are re-weighted according to the information in both the option and spot price data. The method is illustrated using the Heston (1993) stochastic volatility model, based on data simulated to mimic the features of recent S&P500 spot and option price data. The way in which alternative option pricing models can be ranked, via Bayes Factors and via fit, predictive and hedging performance, is demonstrated.

Keywords: Option Pricing; Stochastic Volatility; Volatility Risk; Bayesian Implicit Inference; Markov Chain Monte Carlo;

JEL Classifications: C11, G13.

*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.

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 which assume the stochastic volatility process of Heston (1993) are reweighted according to the information about volatility which comes from both spot and option price data. In addition to demonstrating the production of marginal posteriors for the unknown elements of the Heston (1993) model, methods for comparing the Heston model with models which make different assumptions about the volatility of returns are presented. These methods involve the construction of Bayes Factors as well as fit, predictive and hedging error densities. The paper also outlines how the assumption of white noise option pricing errors, adopted initially for the sake of simplicity, can be readily extended to incorporate heteroscedasticity and autocorrelation in the option pricing errors, as well as regime shifts across different contract groups.

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 which attempts to draw 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 model of Black and Scholes (1973). 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 basic dataset to be used in the numerical demonstration of the method. The option prices, spot prices and volatilities are all simulated artificially, based on the assumption that the underlying returns process follows the Heston model. In order to render the artificial data as realistic as possible, the simulation process takes into account certain features of recently observed data on the S&P500 index. The numerical results suggest that the methodology does well in picking up the features of the data. Section 5 expands the distributional assumptions invoked for the option pricing errors to accommodate a range of features both across time and moneyness. We provide some concluding comments in Section 6.

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:

$$dS(t) = rS(t)dt + \sqrt{v(t)}S(t)d\varepsilon_1(t) \qquad (1)$$

and

$$dv(t) = \kappa[\theta - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t), \qquad (2)$$

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:

$$dS(t) = \mu S(t)dt + \sqrt{v(t)}S(t)d\varepsilon_1(t) \qquad (3)$$

and

$$dv(t) = \kappa^a[\theta^a - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t), \qquad (4)$$

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

$$\theta = \frac{\kappa^a\theta^a}{\kappa^a + \lambda}, \qquad (5)$$

$$\kappa = \kappa^a + \lambda. \qquad (6)$$

The parameter λ denotes the price of volatility risk, with the assumption that the volatility risk premium, $\lambda(S(t), v(t))$, is proportional to $v(t)$, that is,

$$\lambda(S(t), v(t)) = \lambda v(t), \qquad (7)$$

being invoked. 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 this underlying asset as:

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

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 which 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 the parameters which 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 which 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 2000), Bakshi, Cao and Chen (1997) and Pan (2000) apply classical methods to an extension of the Heston model which accommodates random jumps in the asset price. Of these latter three 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 which 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. Denoting by $\omega^a = \{\kappa^a, \theta^a, \sigma_v\}$, the set of parameters which characterize the actual volatility process, the MH algorithms for v and ω^a respectively use candidate densities which are based on the assumed volatility model in (4), with the algorithm reweighting these draws according to the compatibility of the draws with the option and spot

[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.

price data. Since only the option price data reflects the market's attitude towards volatility risk, only that component of the data set is relevant to inference 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 provide information on the actual mean rate of growth, μ. Both sets of data provide information on the correlation parameter ρ.

There is some similarity between our proposed algorithm and the MCMC algorithm proposed independently by Eraker (2001), 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 a vector, v, rather than performing iterative simulation of each individual variance, $v(t)$, conditional on the remaining the variances. The vector of parameters describing the actual volatility process, ω^a, is also simulated as a block, using an approximation to the structure in (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 possible relevance of Bayesian model averaging in an option pricing context.[4] As the main aim of the paper is methodological, we choose to focus on the Heston model and various close variants, rather than expanding the model to accommodate the jump processes which 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. In this section, we begin by specifying the simplest possible model for the generation of the observed option

[3] Jones (2000) also uses Bayesian methods to draw options-based inferences about a stochastic volatility model. However, his methodology is somewhat different from that proposed here. Martin, Forbes and Martin (2001) use Bayesian methods to estimate alternative nonnormal returns models from option prices.

[4] Although Eraker appears to exploit information in the spot prices in the estimation procedure, it is not quite clear exactly how the spot price process feeds into the MCMC algorithm which he proposes.

[5] Results related to the empirical usefulness of allowing for random jumps in the asset process and/or the volatility process remain rather mixed; see, for example, Eraker (2001).

prices. 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 which affect the ith option price, the option pricing model is initially specified as

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

where u_i is an unobservable pricing error, assumed to have a normal distribution with zero mean and known variance, σ_u^2, N is the number of observed option prices and $q_H(z_i, \phi)$ is the ith theoretical option price as defined in (8). The vector $\phi = \{v, \omega^a, \lambda, \rho\}$ comprises all unobservable elements which characterize q_H. The index i indicates both variation over time and variation across option contracts at a given point in time. 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 (simulated) option prices are observed, with one variance per day, denoted by v_t, being estimated. Hence $v = (v_1, v_2, \ldots, v_n)'$. The structure of the option price panel is assumed to be such that only one spot price is recorded on each day, with a cross-section of option prices, associated with different values for K_i and τ_i, then associated with that same spot price. Hence, the symbol S_i denotes the spot price, S_t, associated with day t on which the ith option price is observed, with there being an $n-$dimensional vector $s = (S_1, S_2, \ldots, S_n)'$ of spot prices on which the N option prices depend. The same situation prevails for the interest rate r_i, there being one observation per day, r_t, and an $n-$dimensional vector of interest rates on which the option prices depend. It is notationally convenient, however, to continue to index all observable factors which 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

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

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 which 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. In Section 5, we generalize beyond (10) by allowing for unknown intercept, slope and variance parameters, heteroscedasticity and autocorrelation in u_i, as well as shifts in the relationship between C_i and $q_H(z_i, \phi)$ across contract groups.[6]

[6] To rule out arbitrage, the distribution of C_i should strictly speaking be truncated from below at $lb_i = \max\{0, S_i - e^{-r_i\tau_i}K_i\}$; see Hull (2000). However, the incorporation of this truncation in the likelihood function means that when (9) is expanded to accommodate the distributional features just alluded to, the additional parameters thereby introduced cannot be as readily incorporated into the numerical scheme. Since experimentation has established that the truncation has only minimal impact on inferences regarding the underlying returns model, 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.

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

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

As already noted, in order to produce simultaneous inference about the parameters of both the risk-neutral and objective processes or, alternatively, 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 the 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.[7]. Suppressing the dependence of $p(c|z,\phi)$ on all elements of z other than the spot price vector s and defining the vector $\delta = \{\phi, \mu\}$ as the full set of unknowns in the problem, the joint posterior density for δ is thus specified as

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

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

$$p(\omega^a, \rho, \lambda, \mu) = p(\omega^a) \times p(\rho) \times p(\lambda) \times p(\mu). \quad (13)$$

That is, the assumption is invoked that the set of parameters which 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. [8]

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 five groups as follows:

1. v
2. ω^a

[7] See Bates (2000) for more on this issue.

[8] 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.

3. λ

4. ρ

5. μ

Starting values are chosen, and a Gibbs-based MCMC algorithm is then applied to produce successive draws of v, ω^a, λ, ρ and μ via the respective conditional posteriors:

1. $p(v|\omega^a, \lambda, \rho, \mu, c, s)$

2. $p(\omega^a|v, \lambda, \rho, \mu, c, s)$

3. $p(\lambda|v, \omega^a, \rho, \mu, c, s)$

4. $p(\rho|v, \omega^a, \lambda, \mu, c, s)$

5. $p(\mu|v, \omega^a, \lambda, \rho, c, s)$

All of these conditionals, apart from that of μ, are nonstandard and, hence, require numerical treatment to produce draws. Given the satisfaction of the relevant regularity conditions (see Tierney, 1994), these draws converge in distribution to a sample from the full joint posterior distribution. We consider the five conditionals in order.

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

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

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

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 decomposition in (14) motivates the following MH algorithm for selecting a draw from $p(v|\omega^a, \lambda, \rho, \mu, c, s)$.

Step 1 For given values for the elements of ω^a, generate the vector of simulated variances, v^*, from the candidate density,

$$p_c(v|\omega^a, \lambda, \rho, \mu, c, s) \propto p(v|\omega^a).$$

In other words, the assumed dynamic process for the variances, as given by (4), is used as the candidate density. We generate v using an Euler discretization of (4):

$$\begin{aligned} v_t &= v_{t-1} + \kappa^a(\theta^a - v_{t-1})\Delta t + \sigma_v\sqrt{v_{t-1}}\sqrt{\Delta t}\varepsilon_{2t} \\ &= \kappa^a\theta^a\Delta t + (1 - \kappa^a\Delta t)v_{t-1} + \sigma_v\sqrt{v_{t-1}}\sqrt{\Delta t}\varepsilon_{2t} \qquad (15) \\ \varepsilon_{2t} &\sim N(0,1); \quad 2, 3, \ldots n, \end{aligned}$$

with the initial value set equal to the long-run mean, i.e. $v_1 = \theta^a$. 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).[9]

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

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

where v^s indicates a starting value for the MH subchain. That is, the MH subchain involves assessing the likelihood of the observed option price vector, c, given the simulated variance vector, v^*, relative to the likelihood of c, given a previous set of variances, v^s, with this ratio of relative likelihoods multiplied by the corresponding relative likelihoods of the observed spot price vector, s. The form of $p(s|v^*, \omega^a, \rho, \mu)$ is determined via an Euler discretization of (3), namely

$$\begin{aligned} \ln S_t &= \ln S_{t-1} + (\mu - 0.5 v_{t-1})\Delta t + \sqrt{v_{t-1}}\sqrt{\Delta t}\,\varepsilon_{1t}, \\ \varepsilon_{1t} &\sim N(0,1); \quad 2, 3, \ldots n, \end{aligned} \tag{16}$$

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

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

where

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

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

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

[9] See Eraker, Johannes and Polson (2001) for evidence that the daily interval is small enough to render the discretization bias arising from (15) negligible.

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

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

$$\begin{aligned} p(\omega^a|v,\lambda,\rho,\mu,c,s) &\propto p\left(c|s,v,\omega^a,\lambda,\rho\right)\times p(s|v,\omega^a,\rho,\mu)\times p(v|\omega^a)\times p(\omega^a) \\ &\propto p\left(c|s,v,\omega^a,\lambda,\rho\right)\times p(s|v,\omega^a,\rho,\mu)\times p(\omega^a|v). \end{aligned} \tag{19}$$

From (19) 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:

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

Further defining

$$\zeta(L) = 1-\zeta L, \tag{21}$$

$$y_t = \frac{\zeta(L)v_t}{\sqrt{v_{t-1}}\sqrt{\Delta t}}, \tag{22}$$

$$x_t = \frac{1-\zeta}{\sqrt{v_{t-1}}\sqrt{\Delta t}} \tag{23}$$

and

$$e_t = \sigma_v\varepsilon_{2t},$$

where L is the lag operator with respect to time period Δt, (20) can be written as

$$\begin{aligned} y_t &= \theta^a x_t + e_t, \\ e_t &\sim N(0,\sigma_v^2), \\ t &= 2,3,\ldots,n. \end{aligned} \tag{24}$$

Given the form of (24), the natural decomposition of $p(\omega^a|v)$ is

$$p(\omega^a|v) = p(\theta^a|\kappa^a, \sigma_v, v) \times p(\sigma_v|\kappa^a, v) \times p(\kappa^a|v). \tag{25}$$

We adopt a prior of the form

$$\begin{aligned} p(\omega^a) &= p(\kappa^a)p(\theta^a)p(\sigma_v) \\ &\propto \frac{1}{\sigma_v} \times 1_{(\kappa^a>0)} \times 1_{(\theta^a>0)} \times 1_{(2\kappa^a\theta^a>\sigma_v^2)}, \end{aligned} \tag{26}$$

where the two indicator functions, $1_{(\kappa^a>0)}$ and $1_{(\theta^a>0)}$, restrict κ^a and θ^a respectively to the positive region[10] and the third indicator function, $1_{(2\kappa^a\theta^a>\sigma_v^2)}$, ensures that the variances associated with (4) are always positive. Suppressing the indicator functions, the prior in (26) implies that the first two component densities in (25) are given by

$$\theta^a|\kappa^a, \sigma_v, v \sim N(\widehat{\theta}^a, \sigma^2_{\theta^a}), \tag{27}$$

and

$$\sigma_v|\kappa^a, v \sim IG(v, s_v^2), \tag{28}$$

with

$$\begin{aligned} \widehat{\theta}^a &= \sum_{t=2}^{n}(x_t y_t) / \sum_{t=2}^{n}(x_t^2) \\ \sigma^2_{\theta^a} &= \sigma_v^2 / \sum_{t=2}^{n}(x_t^2), \\ s_v^2 &= \sum_{t=2}^{n}(y_t - \widehat{\theta}^a x_t)^2/\nu \end{aligned}$$

and $\nu = n - 2$.[11] It is readily shown that the third component density has the form

$$p(\kappa^a|v) \propto \left(\sum_{t=2}^{n}(y_t - \widehat{\theta}^a x_t)^2\right)^{-(n-2)/2} \times \left(\sum_{t=2}^{n}(x_t^2)\right)^{-1/2}. \tag{29}$$

The first term in (29) can, in turn, be re-expressed as

$$\left(\sum_{t=2}^{n}(y_t - \widehat{\theta}^a x_t)^2\right)^{-(n-2)/2} \propto \left(1 + \frac{1}{\eta\sigma^2_{\kappa^a}}(\kappa^a - \widehat{\kappa}^a)^2\right)^{-(\eta+1)/2}, \tag{30}$$

where

$$\widehat{\kappa}^a = \frac{b}{a}$$

and

$$\sigma^2_{\kappa^a} = (c - \frac{b^2}{a})/(\eta a),$$

[10] 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.

[11] In Section 4 we introduce a non-uniform prior on θ^a. This prior is used in order to further aid in the discrimination between the parameters of the risk-neutral and objective processes respectively. Since the informative prior is specified as a normal density, the basic structure of both $p(\omega^a|v)$ and the simulation algorithm as outlined in this section, remains unchanged.

with

$$
\begin{aligned}
a &= \sum_{t=2}^{n} v_{t-1} - \left((n-1)^2 / \sum_{t=2}^{n} \frac{1}{v_{t-1}}\right), \\
b &= \sum_{t=2}^{n} v_t - \left((n-1) \sum_{t=2}^{n} \frac{v_t}{v_{t-1}} / \sum_{t=2}^{n} \frac{1}{v_{t-1}}\right), \\
c &= \sum_{t=2}^{n} \frac{v_t^2}{v_{t-1}} - \left(\left[\sum_{t=2}^{n} \frac{v_t}{v_{t-1}}\right]^2 / \sum_{t=2}^{n} \frac{1}{v_{t-1}}\right)
\end{aligned}
$$

and $\eta = n-3$. For $\eta > 3$ the function in (30) is proportional to a well-defined Student t density for κ^a, with finite mean, $\widehat{\kappa}^a$, and variance, $\sigma^2_{\kappa^a}$. As such, the following MH algorithm for drawing a value for ω^a from $p(\omega^a|v, \lambda, \rho, \mu, c, s)$ is motivated.

Step 1 For a given vector value for v, generate ω^a from the candidate density:

$$
p_c(\omega^a|v, \lambda, \rho, \mu, c, s) \propto p_c(\theta^a|\kappa^a, \sigma_v, v) \times p_c(\sigma_v|\kappa^a, v) \times p_c(\kappa^a|v), \qquad (31)
$$

where $p_c(\theta^a|\kappa^a, \sigma_v, v)$ denotes the normal density for θ^a described in (27), $p_c(\sigma_v |\kappa^a, v)$ denotes the inverted gamma density for σ_v described in (28) and $p_c(\kappa^a|v)$ denotes the Student t density for κ^a described in (30). In other words, use as candidate density, the 'posterior density' for ω^a given the latent variance vector, v, with the marginal component of this candidate density for κ^a approximated by the Student t density in (30). Values of κ^a, θ^a and σ_v which violate the prior restrictions included in (26) are discarded in the simulation process. Given the decomposition in (25), with $p(\kappa^a|v)$ as given in (30), a draw from $p(\omega^a|v)$ can readily occur via the method of composition.

Step 2 Select the drawn value, ω^{a*}, as a drawing from $p(\omega^a|v, \lambda, \rho, \mu, c, s)$ with probability

$$
\begin{aligned}
\psi &= \min\left\{\frac{p(\omega^{a*}|v, \lambda, \rho, \mu, c, s)}{p_c(\omega^{a*}|v, \lambda, \rho, \mu, c, s)} / \frac{p(\omega^{as}|v, \lambda, \rho, \mu, c, s)}{p_c(\omega^{as}|v, \lambda, \rho, \mu, c, s)}, 1\right\} \\
&= \min\{\frac{p\,(c|s, v, \omega^{a*}, \lambda, \rho) \times p(s|v, \omega^{a*}, \rho, \mu) \times p(\omega^{a*}|v)}{p_c(\omega^{a*}|v, \lambda, \rho, \mu, c, s)} \\
&\quad / \frac{p\,(c|s, v, \omega^{as}, \lambda, \rho) \times p(s|v, \omega^{as}, \rho, \mu) \times p(\omega^{as}|v)}{p_c(\omega^{as}|v, \lambda, \rho, \mu, c, s)}, 1\} \\
&= \min\{\frac{p\,(c|s, v, \omega^{a*}, \lambda, \rho) \times p(s|v, \omega^{a*}, \rho, \mu) \times (\sum_{t=1}^{n}(x_t^{*2}))^{-1/2}}{p\,(c|s, v, \omega^{as}, \lambda, \rho) \times p(s|v, \omega^{as}, \rho, \mu) \times (\sum_{t=1}^{n}(x_t^{s2}))^{-1/2}}, 1\},
\end{aligned}
$$

where ω^{as} indicates a starting value for the MH subchain, x_t^s denotes x_t, as defined in (23), evaluated at ω^{as} and x_t^* denotes x_t evaluated at ω^{a*}. That is, the MH subchain involves assessing the ratio of the relative likelihoods of c given simulated and previous values for the parameters of the variance process, multiplied by the ratio of corresponding likelihoods for the spot price data, modified by an adjustment factor, $\left(\sum_{t=2}^{n}(x_t^{*2})\right)^{-1/2} / \left(\sum_{t=1}^{n}(x_2^{s2})\right)^{-1/2}$, which results from the approximation of $p(\omega^a|v)$ by the candidate density in (31).[12]

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

The parameter λ, being one-dimensional, could be readily simulated from its conditional posterior using the inverse cumulative distribution function (icdf) technique. This would involve specifying the function:

$$p(\lambda|v, \omega^a, \rho, \mu, c, s) \propto p\left(c|s, v, \omega^a, \lambda, \rho\right) \times p(\lambda) \tag{32}$$

at a grid of values for λ, numerically normalizing the cumulative distribution function and drawing a value of λ at random via the inverse of this function.[13] This approach, however, is computationally burdensome, since it involves evaluation of the $(N \times 1)$ vector $q_H(z_i, \phi)$ at each grid point for λ. We adopt instead a computationally efficient MH subchain, 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, 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,

$$q_H(z_i, \lambda|v, \omega^a, \rho) \approx q_H(z_i, \lambda^{\#}|v, \omega^a, \rho) + q_H^{\prime}(z_i, \lambda^{\#}|v, \omega^a, \rho)(\lambda - \lambda^{\#}), \tag{33}$$

where $q_H^{\prime}(z_i, \lambda^{\#}|v, \omega^a, \rho)$ 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

$$q_H^{\prime}(z_i, \lambda^{\#}|v, \omega^a, \rho) \approx \frac{q_H(z_i, (\lambda^{\#} + h)|v, \omega^a, \rho) - q_H(z_i, \lambda^{\#}|v, \omega^a, \rho)}{h} \tag{34}$$

for h small. Substitution of (33) and (34) for $q_H(z_i, \lambda|v, \omega^a, \rho)$ in the expression for $p\left(c|s, v, \omega^a, \lambda, \rho\right)$ in (32) produces a conditional candidate density, $p_c(\lambda|v, \omega^a, \rho, \mu, c, s)$, of the form[14]

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

[12]Following the line of argument adopted in Chao and Phillips (1998), Kleibergen and van Dijk (1994) and Martin (2000, 2001), the approximation to $p(\omega^a|v)$ used as the candidate density can be viewed as equivalent to $p(\omega^a|v)$ when an information matrix based prior of the form $p(\theta^a|\kappa^a) = \left(\sum_{t=2}^{n}(x_t^2)\right)^{1/2}$ is used. This prior can be justified on the grounds of serving to offset the near lack of identification of θ_a which, for small samples, exists in the region $\kappa^a \approx 0$. We do not invoke this prior here since this identification issue does not cause problems for the sample size used in the numerical illustration in Section 4.

[13]It will be noted from (32) that λ enters the analysis only via $p\left(c|s, v, \omega^a, \lambda, \rho\right)$, in which λ affects the value of the theoretical option price q_H, and via the prior on λ, $p(\lambda)$. The value of λ does not affect the generation of s and v in (3) and (4).

[14]For notational consistency, we continue to condition both the candidate density for λ and the conditional density for λ on μ, even though neither is dependent on μ.

where

$$\widehat{\lambda} = \frac{\sum_{i=1}^{N}(C_i - [q_H(z_i, \lambda^{\#}|v, \omega^a, \rho) - q_H^{\theta}(z_i, \lambda^{\#}|v, \omega^a, \rho)\lambda^{\#}]q_H^{\theta}(z_i, \lambda^{\#}|v, \omega^a, \rho)}{\sum_{i=1}^{N}(q_H^{\theta}(z_i, \lambda^{\#}|v, \omega^a, \rho))^2}. \quad (35)$$

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

$$p_c(\lambda|v, \omega^a, \rho, \mu, c, s) \propto \exp(\frac{-1}{2\overline{\overline{var}}(\lambda)}(\lambda - \overline{\overline{\lambda}})^2), \quad (36)$$

with

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

and

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

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 $p_c(\lambda|v, \omega^a, \rho, \mu, c, s)$ and chosen with probability,

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

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

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

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

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

$p\left(c|s,v,\omega^a,\lambda,\rho\right)$, the first component of the candidate is defined as a normal density for ρ, with mean and variance given respectively by

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

and

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

where $q_H^{\theta}(z_i,\rho^{\#}|v,\omega^a,\lambda)$ 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

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

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

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

and

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

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

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

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

Using the expressions in (17) and (18), it follows that

$$\mu|v, \omega^a, \rho, s \sim N(\widehat{\mu}, \sigma_\mu^2),$$

where

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

and

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

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.

Implementation of the hybrid 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 which differ significantly from zero have been reported (see Guo, 1998 and Eraker, 2001, amongst others). An assessment of this restricted model thus amounts to an assessment of the attitude to volatility risk which 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 regularity of negative skewness in returns, an assessment of this restricted model corresponds to an assessment of whether or not returns are skewed and/or whether or not option prices have factored in skewed returns.

Finally, the restriction $\sigma_v = 0$ equates to the assumption of constant volatility. Given the assumption of normal returns, this restriction equates to the assumption of the 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 which 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. $\delta_1 = \{v, \omega^a, \lambda, \rho, \mu\}$
2. $\delta_2 = \{v, \omega^a, \rho, \mu\}$
3. $\delta_3 = \{v, \omega^a, \lambda, \mu\}$
4. $\delta_4 = \{\theta^a, \mu\}$

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,

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

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

$$p(c, s|M_k) = \int_{\delta_k} p(c|s, \delta_k, M_k)\, p(s|\delta_k, M_k) p(\delta_k|M_k) d\delta_k; \quad k = 1, 2, \ldots 4. \tag{37}$$

In the expression for $p(c, s|M_k)$ in (37), δ_k denotes the vector of unobservables which characterize model M, $p(c|s, \delta_k, M_k)$ 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 (37) 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:

$$SD_{21} = \frac{p(\lambda = 0|M_1, c, s)}{p(\lambda = 0|M_1)}, \tag{38}$$

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 (38) to be an equivalent representation of the Bayes Factor for M_2 versus M_1, BF_{21}, it must hold that

$$p(\delta_2|M_1, \lambda = 0) = p(\delta_2|M_2), \tag{39}$$

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 (39) can be interpreted as the requirement that the 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

$$SD_{31} = \frac{p(\rho = 0|M_1, c, s)}{p(\rho = 0|M_1)}. \tag{40}$$

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

$$p(\theta^a, \mu | M_1, \sigma_v = 0) = p(\theta^a, \mu | M_4), \tag{41}$$

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 (38) and (40), namely,

$$SD_{41} = \frac{p(\sigma_v = 0 | M_1, c, s)}{p(\sigma_v = 0 | M_1)}. \tag{42}$$

Since (41) 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. Assuming that, for computational convenience, we were to specify a conjugate inverted gamma prior for σ_v, this prior would have a limit of zero when $\sigma_v = 0$, thereby producing a zero value for the denominator in (42). If the posterior for σ_v under the Heston model were concentrated near zero, the numerator may be nonzero at $\sigma_v = 0$. In this case, SD_{41} would approach infinity and hence infinitely favour M_4. Since all derivatives of the denominator approach zero for $\sigma_v = 0$, application of l'Hopital's rule would produce either an indeterminate result or continue to produce an infinite favouring of M_4. Moreover, the calculation of such a limit, requiring numerical differentiation of the numerator in (42), with that numerator itself estimated numerically, is likely to be highly unreliable. 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

$$res_i = C_i - q(z_i, \phi_k), \quad i = 1, 2, \ldots, N, \tag{43}$$

where $q(z_i, \phi_k)$ denotes the theoretical option price associated with model M_k, $k = 1, 2, \ldots 4$ and $\delta_k = \{\phi_k, \mu\}$. 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

$$q_{BS}(z_i, \phi_4 = \theta^a) = S_i \Phi(d_1) - K_i e^{-r_i \tau_i} \Phi(d_2), \tag{44}$$

where

$$d_1 = \frac{\ln(S_i/K_i) + (r_i + \theta^a/2)\tau_i}{\sqrt{\theta^a \tau_i}}, \tag{45}$$

$$d_2 = \frac{\ln(S_i/K_i) + (r_i - \theta^a/2)\tau_i}{\sqrt{\theta^a \tau_i}} \tag{46}$$

and $\Phi(x)$ denotes the cumulative normal distribution function evaluated at x; see, for example, Hull (2000). To reduce the computational burden associated with estimating the posterior distribution of res_i, we assume that $p(res_i|c, s)$ is approximately normal. Given this assumption, an approximate 95% Highest Posterior Density (HPD) interval for res_i is given by

$$INT_i = \overline{res}_i \pm 1.96 \times sd(res_i), \tag{47}$$

where $\overline{res}$ and $sd(res_i)$ are respectively the mean and standard deviation of res_i values associated with the simulated values for ϕ_k produced by the relevant sampling algorithm. For M_1, the 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 $\delta_4 = \{\theta^a, \mu\}$ are generated from the posterior density for δ_4, given by

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

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

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

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

$$\begin{aligned} p(\ln s|\theta^a, \mu) &= (2\pi)^{-n/2}(\theta^a\Delta t)^{-n/2} \times \\ &\quad \exp\left\{(-1/[2\theta^a\Delta t]) \sum_{t=2}^{n}(\ln S_t - [\ln S_{t-1} + (\mu - 0.5\theta^a)\Delta t])^2\right\}. \end{aligned} \tag{48}$$

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 INT_i intervals which cover zero can be calculated for each model M_k, with the best fitting model being the one for which this proportion is the highest. If the observed option prices used to define (43) 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:

$$p(C_f|c,s) = \int_{\phi_k} p(C_f|c,s,\phi_k)p(\phi_k|c,s)d\phi_k, \tag{49}$$

where $p(C_f|c,s,\phi_k)$ is a normal density with known variance σ_u^2. In order to impose the lower bound which C_f must exceed in order for arbitrage opportunities to be avoided, in the construction of (49) we specify $p(C_f|c,s,\phi_k)$ as the **truncated** normal density

$$p(C_f|c,s,\phi_k) = \frac{(2\pi\sigma_u^2)^{-1/2}}{(1-\Phi(slb_f))}\exp\left(-\frac{1}{2\sigma_u^2}[C_f - q(z_f,\phi_k)]^2\right), \tag{50}$$

where

$$slb_f = \frac{lb_f - q(z_f,\phi_k)}{\sigma_u}$$

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 $p(\phi_k|c,s)$, $\phi_k^{(j)}$, $j = 1,2,\ldots,B$, can be used to construct an estimate of $p(C_f|c,s)$ as

$$\widehat{p(C_f|c,s)} = \frac{1}{B}\sum_{j=1}^{B} p(C_f|c,s,\phi_k^{(j)}). \tag{51}$$

In Section 4, prediction intervals constructed from (51) are used to rank the predictive performance of the models.[16]

3.3.3 Hedging Performance

An important measure of the performance of the alternative stochastic 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

$$C_t - N_k S_t, \tag{52}$$

[16]Note that only ϕ_k is relevant in the assessment of the fit and predictive performance of model M_k, as the remaining parameter, μ, of which $\delta_k = \{\phi_k, \mu\}$ is comprised, does not impact directly on the option pricing components of the model. The same point holds for the hedging criterion discussed below.

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

$$\frac{\partial q_H}{\partial v_t} = S_t \frac{\partial P_1}{\partial v_t} - K e^{-r_t \tau} \frac{\partial P_2}{\partial v_t}, \tag{53}$$

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,

$$H_1 = (C_t - N_1 S_t)\, e^{r_t \tau} - (C_{t+1} - N_1 S_{t+1}), \tag{54}$$

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 $p(H_1|c, s)$[17]. 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 (45), renders the portfolio in (52) delta-hedged; see Hull (2000). For the three stochastic volatility models, the derivatives with respect to v_t which enter (53) 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 could

[17] See Footnote 16.

be used to produce a model-averaged predictive density, which could, in turn, be used as the tool for prediction rather than the predictive associated with any one particular model.[18] 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:

$$p_a(C_f|c,s) = \sum_{k=1}^{3} p(C_f|M_k, c, s)P(M_k|c,s), \tag{55}$$

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

4 Numerical Illustration: Simulated S&P 500 Option Prices

4.1 Data Description

The methodology is demonstrated using option and spot price data artificially generated to reflect the Heston model. Spot prices and variances are first generated assuming the bivariate process in (3) and (4). In order to produce empirically relevant numerical results, we simulate spot price data which has some of the features of recent observations on the S&P500 index. Specifically, we simulate daily index values across 1000 trading days, starting with the index observed on the first trading day of 1998, $S_1 = 97504$. The parameters $\kappa^a$and θ^a of the actual variance process are chosen to reflect the results of a preliminary volatility analysis of actual returns on the S&P500 index over the 1998 to 2001 period.[19] This preliminary analysis involves the estimation of a Generalized Autoregressive Conditional Heteroscedastic (GARCH) model for returns, with the value of κ^a then chosen to match the value of the persistence measure in the GARCH model and θ^a chosen to reflect the long-run mean implied by the GARCH parameter estimates. The "volatility of volatility" parameter, σ_v, and the correlation parameter, ρ, are chosen in such a way that the implied degree of kurtosis and skewness respectively bears some resemblance to that evident in the unconditional distribution of S&P500 returns over the 1998 to 2001 period.

This part of the data generating procedure produces 1,000 spot prices and 1,000 variances, one for each assumed trading day. On each trading day, it is further assumed that 20 prices of option contracts which vary according to the strike price, K, are observed. The theoretical prices of these contracts are simulated by substituting into (8) the assumed strike prices, the spot price, S_t, and variance, v_t, associated with that day, along with the chosen values of the parameters. The interest rate used in

[18] See Geweke (1999) for discussion of the principles of Bayesian model averaging.

[19] Daily observations on the S&P500 index were obtained from Standard and Poors. The preliminary returns analysis was performed using the EVIEWS software package.

(8) is the average of daily observations on the US three-month bill rate, from the 1st of January, 1998 to the 20th of December, 2001, $r = 0.05$.[20] The strike prices are chosen so as to imply a range of 'option moneyness' from 0.8 to 1.6, where moneyness is defined as the ratio of S_t to K. Since the S&P500 index pays dividends, S_t in (8) is replaced by $S_t e^{-d\tau}$, where d in the actual annualized rate of dividend payment on the S&P500 index over the 1998 to 2001 period and τ is the maturity of the option.[21] We set $\tau = 30/365$ for all options, in order to simulate a sample of short-term option prices. Given the sample theoretical Heston option prices, $q_H(z_i, \phi_1)$, $i = 1, 2, \ldots, 20,000$, the sample of option prices to be used as the option price data set in the analysis, is produced via (9), with σ_u set to 80 (cents). The final sample of option prices is then filtered by discarding all prices which violate the Merton no-arbitrage lower bound, $lb = \max\{0, S_t e^{-d\tau} - e^{-r\tau} K\}$. The out-of-sample assessments of the alternative volatility models are based on a further 10 days of simulated data.

The values assumed for the parameters of the Heston model are as follows: $\kappa^a = 2.5$; $\theta^a = 0.04$; $\sigma_v = 0.05$; $\lambda = -1$; and $\rho = -0.1$. The value specified for the mean reversion parameter translates into daily persistence parameters for a $GARCH(1,1)$ model of $1 - 2.5/252 = 0.99$. This value, as mentioned above, is reflective of the volatility persistence discovered in the preliminary analysis of recent $S\&P500$ returns data. Similarly, the long-run mean variance of $\theta^a = 0.04$, corresponds approximately to the long-run mean variance estimate produced by the same preliminary $GARCH(1,1)$ analysis. The value of 0.04 for θ^a implies an annualized mean volatility value of 20%. The negative value specified for lambda is consistent with the empirical literature in the area.[22] The value of -1 for λ implies that the risk-neutral variance process has a long run mean of $\theta = 0.067$ with reversion parameter $\kappa = 1.5$. The value of $\rho = -0.1$ produces an unconditional distribution for returns with slight negative skewness, again a feature which tallies with the corresponding feature in actual returns on the index. The value of $\sigma_v = 0.05$, however, does not produce a degree of excess kurtosis in returns which matches that in actual returns. A value of approximately 0.3 is required to achieve this match, a value which causes problems for the numerical scheme. Realism is thereby sacrificed at this point for the sake of computational ease. A value of 0.08 is specified for μ, reflecting a situation in which $\mu > r$. Note however that this value for μ is almost twice the size of the actual mean annual return on the S&P500 index over the period used for the preliminary analysis.

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 volatilities associated with the Heston model. All results are produced by running the MCMC algorithm for 1100 iterations, with the first 100 iterations discarded. The starting values for the parameters are as follows: $\kappa^a = 3.0$; $\theta^a = 0.04$; $\sigma_v = 0.06$;

[20] The daily interest rate observations were extracted from Datastream.

[21] Daily dividend data for this period was obtained from Standard and Poors.

[22] A more detailed interpretation of this parameter is provided in the appendix.

$\lambda = -0.5$; $\rho = -0.3$; and $\mu = 0.06$. The starting vector value for v is produced by simulating a random vector of variances from the process in (4) with the parameters set at their starting values. 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.[23] A very tight prior is imposed on θ^a, specifically a normal distribution with mean of 0.04 and standard deviation of 0.0001. The prior mean value of θ^a corresponds to the "true" value of θ^a which underlies the data, as well as corresponding to the starting value of θ^a in the MCMC chain. Such specifications associated with θ^a can be justified as follows. With the true value of θ^a having been specified on the basis of a preliminary analysis of returns data, both the prior and the starting value for the θ^a chain can be viewed as reflecting the results of this preliminary analysis. This is indeed a feasible approach to adopt when dealing with actual empirical data and, from a computational point of view, serves to produce a more accurate estimate of λ in particular.[24]

The acceptance rates for the four MH subchains vary considerably. The subchains for the parameters λ and ρ have acceptance rates of approximately 10% and 91% respectively. The acceptance rate for the vector ω^a is about 20% and that of the vector v only 0.5%. With regard to the vector v, most of the accepted iterates occur early on in the chain, with the individual v_t converging quickly to values close to the true long-run mean underlying the data, that is, $\theta^a = 0.04$. This is to be expected given the very informative prior information specified for both θ^a and the long-run mean of the vector v. The normal candidate densities for both λ and ρ, constructed as described in Sections 3.2.3 and 3.2.4 respectively, 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. Since v is updated only rarely, this represents a considerable computational saving.

Table 1 reports the mean, mode and (approximate) Highest Posterior Density (HPD) estimates for the parameters and selected volatilities, with the latter presented as the square root of the corresponding element of v[25]. As would be expected, given the very tight prior specified for θ^a around a mean equal to the true value, the posterior estimates of θ^a are very close to the true value of 0.04. The estimates of κ^a are slightly larger than the true value of 2.5. However, remembering that a true value of $\kappa^a = 2.5$ corresponds to a daily persistence measure of 0.99, the point estimates of approximately 5.0 reported in Table 1 translate into persistence estimates of approx-

[23] 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 which fall beyond them.

[24] Strictly speaking, the inclusion of both spot price and option price data is sufficient to identify κ^a, θ^a and λ. However, with computational issues precluding the use of a very long MCMC chain, the imposition of very tight and accurate information about one of these three parameters has been found to speed up convergence of the chain to the correct region of the parameter space, for the remaining two parameters.

[25] An HPD interval is one with the specified probability coverage, whose inner ordinates are not exceed 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.

Table 1: Marginal Posterior Density Results for the Heston Model

	Mode	Mean	95% HPD interval
Parameter			
κ^a	4.1000	5.1405	(2.8000, 8.6000)
θ^a	0.0400	0.0399	(0.0399, 0.0400)
σ_v	0.0440	0.0439	(0.0400, 0.0460)
λ	-2.6000	-2.4980	(-2.8000, -2.0000)
ρ	-0.0200	-0.0170	(-0.0800, 0.0400)
μ	0.1500	0.1202	(-0.0800, 0.3300)
Volatility			
$\sqrt{v(10)}$	0.2050	0.2012	(0.1920, 0.2070)
$\sqrt{v(100)}$	0.2110	0.2058	(0.1960, 0.2140)
$\sqrt{v(500)}$	0.1970	0.2004	(0.1950, 0.2120)
$\sqrt{v(1000)}$	0.2030	0.2026	(0.2020, 0.2100)

imately 0.98, with the interval estimate translating into a (daily) persistence interval of (0.965, 0.999). All estimates are thus well within the realm of typical estimates of such measures. The point estimates of lamda, although more negative than the true value of -1.0, are also reasonably accurate when viewed on a daily scale. The interval estimate covers only values in the negative region. Note also that the estimates for θ^a, κ^a and λ, when taken together imply estimates for the risk-neutral parameters, θ and κ, which are quite close to the true values. For example, posterior mean estimates of θ^a, κ^a and λ, imply estimates of 0.07 and 2.64 for θ and κ respectively.

The point estimates of the mean rate of return on S_t, μ, are higher than the true value of 0.08, with the interval estimate being rather wide. The latter does however cover the true value. The estimates of ρ are also too large. In this case, the interval estimate does not cover the true value of -0.1. As will be illustrated below, this inaccuracy in the estimation of ρ is associated with a Bayes Factor for $M_3 : \rho = 0$ which (incorrectly) favours M_3 over M_1. The estimates of σ_v are reasonably accurate although, again, the interval estimate does not cover the true value.

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

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

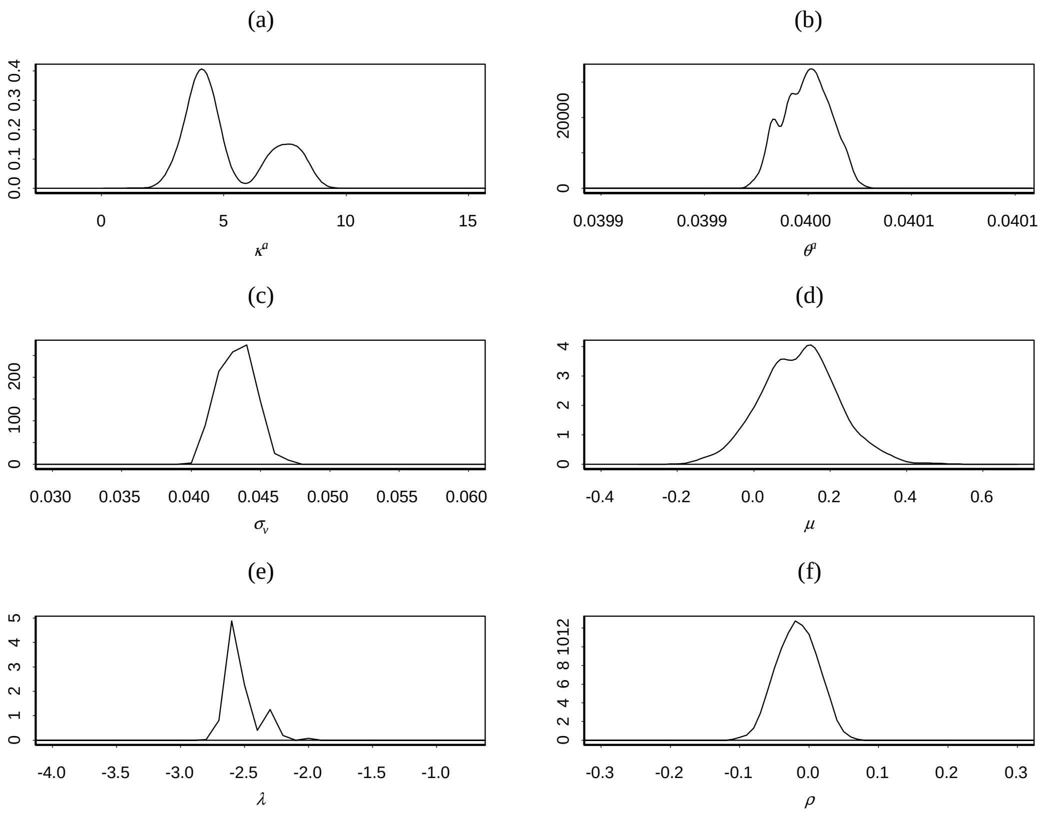
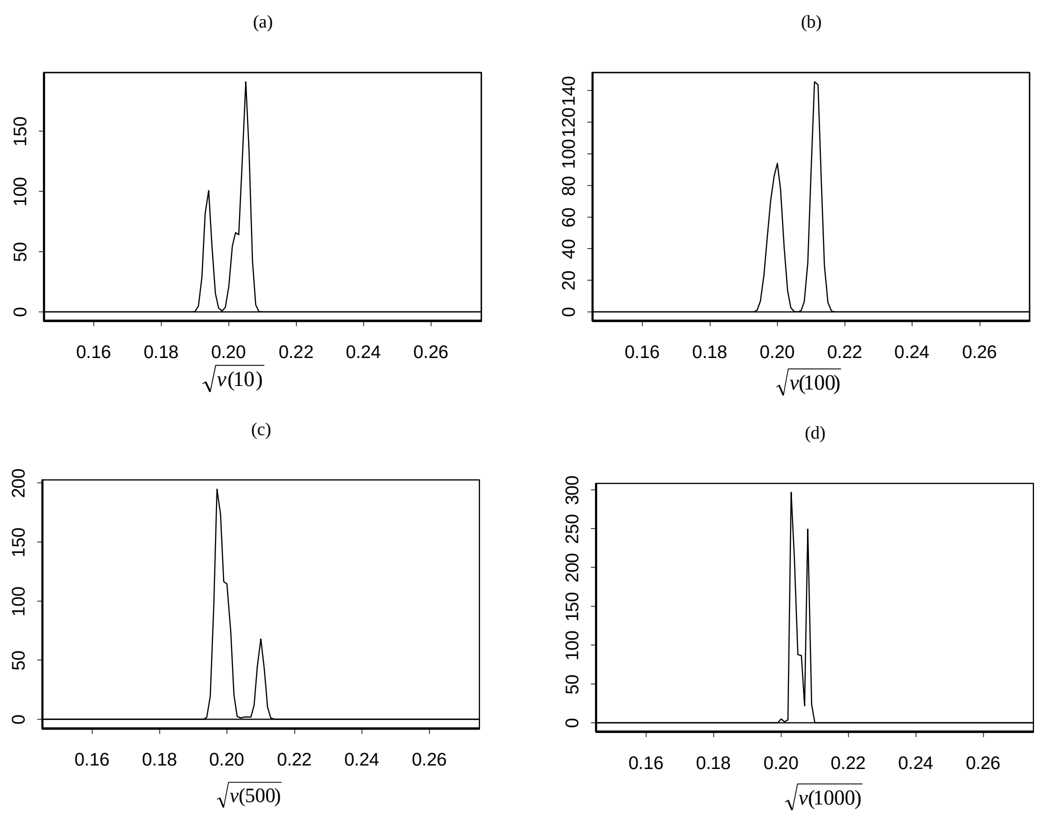


Figure 2: Marginal Posterior Densities for Selected Volatilities from the Heston Model.



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. We do not invoke the model averaging idea here, since the data has been explicitly generated from a single model, M_1, rather than from a variety of different models.

From the interval estimates reported in Table 1, it is clear that two of the three restricted models, M_2 and M_4, are rejected by the data, with neither of the intervals covering the relevant parameter restrictions. The interval estimate for ρ however, does cover the value of $\rho = 0$, thereby lending support to this model. 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 $\sqrt{\theta^a}$ which is extremely concentrated around a single value, namely $\sqrt{\theta^a} = 0.1982$. With only this parameter affecting the theoretical option prices, we report the predictive and hedging results conditional on this single value. A fit assessment is not made since this is based an (approximate) HPD interval for the residual quantities and such an interval would, in this case, have virtually zero width around the central 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_3, with virtually zero posterior weight assigned to both M_2 and the correct model M_1. This result would appear to reflect the somewhat inaccurate (point and interval) estimates of ρ reported above, with more posterior weight assigned to the region around $\rho = 0$ than is appropriate, given the nature of the true data.

The fit and predictive results reported in Table 3, however, rank M_3 below the true model, M_1. The results represent the proportion of times that each criterion is satisfied for all out of sample observations. With 10 days of out-of-sample option prices generated, associated with 20 different strike prices on each day, these results are proportions of 200. According to the fit criterion, the models are ranked in the order M_2, M_1, M_3 and M_4. The predictive criteria only marginally favour M_2, with there being negligible difference between the predictive performances of M_1 and M_2. Again the models are ranked in the order M_2, M_1, M_3 and M_4. The BS model, as based on the single constant volatility estimate of 0.1982, has notably inferior predictive performance relative to all of the other three models.

In Table 4 the hedging error results associated with the four models are reported. Hedging errors are calculated using (54) for M_1 and the version of (54) appropriate for the remaining models, as described in the text immediately following (54). Errors are calculated for one day, five days and ten days ahead, with the portfolio constructed at

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	1.0000	1.5334×10^{-304}	10.6460
M_2		1.0000	6.9427×10^{-304}
M_3			1.0000
Prob$(M_k \mid c)$	0.08587	1.3167×10^{-305}	0.91413

Table 3: Fit and Predictive Performance Measures

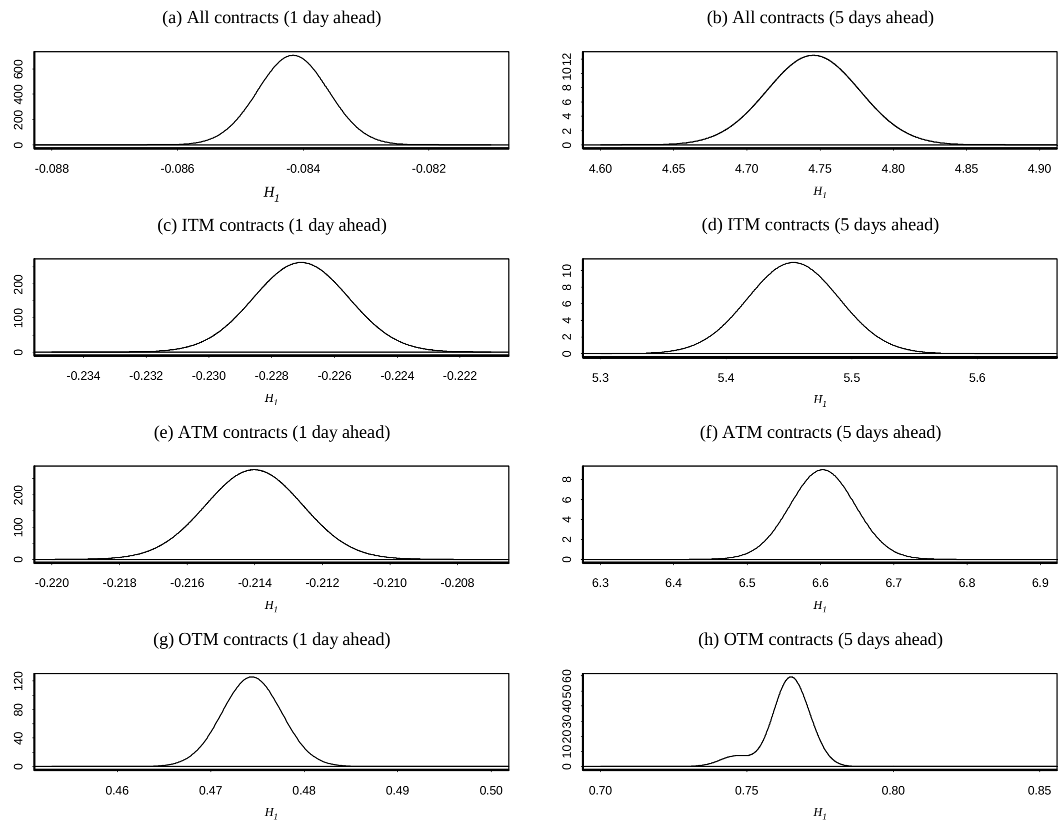
Criterion[(a)]	M_1	M_2	M_3	M_4
Zero in 95% Residual Interval	0.2000	0.2850	0.1450	n.a[(b)]
C_f in Interquartile Predictive Interval[(c)]	0.4000	0.4100	0.3700	0.1500
C_f in 95% Predictive Interval	0.6650	0.6700	0.6450	0.5400

(a) All figures represent proportions of 200.

(b) Since only a single point estimate of the BS volatility has been produced, this interval cannot be constructed.

(c) The 95% Interval is the interval which excludes 2.5% in the lower and upper tails of the predictive distribution. This interval equals the 95% HPD interval only for those predictives which are symmetric around a single mode.

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



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 20 contracts are calculated. The errors are then averaged across all contracts and across respectively, the out-of-the money, at-the-money and in-the-money contracts.[26] It is these averaged figures to which the summary statistics in Table 4 relate and whose posterior densities are graphed in Figure 3.

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 0.001% and 0.02% 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. By the time the tenth out-of-sample day has been reached, the lack of rebalancing produces hedging errors which are unacceptable for all models.

[26]These contract categories correspond respectively to the moneyness ranges: $(S/K) < 0.96$; $0.96 \leq S/K < 1.12$; $1.12 \leq S/K$.

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

Moneyness (S/K)	M_1	M_2	M_3	$M_4^{(a)}$
		One Day Ahead		
All	-0.0841	-0.0841	-0.0841	-0.0842
	(-0.0853, -0.0831)	(-0.0853, -0.0831)	(-0.0853, -0.0831)	n.a.
ITM	-0.2269	-0.2269	-0.2269	-0.2272
	(-0.2300, -0.2241)	(-0.2300, -0.2241)	(-0.2301, -0.2241)	n.a.
ATM	-0.2138	-0.2138	-0.2139	-0.2146
	(-0.2168, -0.2112)	(-0.2168, -0.2112)	(-0.2169, -0.2113)	n.a.
OTM	0.4740	0.4739	0.4741	0.4753
	(0.4681, 0.4806)	(0.4681, 0.4806)	(0.4682, 0.4807)	n.a.
		Five Days Ahead		
All	4.7411	4.7412	4.7413	4.7456
	(4.6831, 4.8080)	(4.6833, 4.8082)	(4.6834, 4.8082)	n.a.
ITM	5.4485	5.4483	5.4494	5.4595
	(5.3815, 5.5255)	(5.3815, 5.5250)	(5.3825, 5.5260)	n.a.
ATM	6.5969	6.5966	6.6007	6.6270
	(6.5190, 6.6940)	(6.5150, 6.6900)	(6.5200, 6.6940)	n.a.
OTM	0.7629	0.7646	0.7573	0.7226
	(0.7424, 0.7773)	(0.7274, 0.7791)	(0.7471, 0.7683)	n.a.
		Ten Days Ahead		
All	60.6208	60.6231	60.6242	60.6794
	(59.8800, 61.4760)	(59.8820, 61.4780)	(59.8840, 61.4800)	n.a.
ITM	78.0185	78.0151	78.0334	78.1834
	(77.0660, 79.1200)	(77.0600, 79.1180)	(77.0800, 79.1360)	n.a.
ATM	64.6960	64.6903	64.7554	65.1316
	(63.9340, 65.6610)	(63.8830, 65.6370)	(63.9640, 65.6710)	n.a.
OTM	4.3524	4.3796	4.2657	3.7153
	(4.0570, 4.5040)	(3.8010, 4.5200)	(4.2010, 4.3330)	n.a.

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

In Figure 3 the posterior densities for the one day ahead and five days ahead hedging errors associated with M_1 are presented. The densities for models M_2 and M_3 are not presented due to the fact that they are so similar to those of M_1. Densities are not constructed for M_4 since the hedging errors are calculated only for a single value of the volatility parameter.

5 Extensions to the Model for the Option Pricing Errors

In this section we illustrate how the simplistic assumptions adopted so far of independent option pricing errors with known variance σ_u^2, can be replaced by a set of more realistic assumptions. As well as outlining the latter, we indicate the implications which they have for the sampling scheme proposed in Section 3.2. The discussion is thus couched within the context of estimating the parameters of the Heston model, with the symbol ϕ being used to represent the vector of unknowns which determine the theoretical option price for that model. However, the discussion applies equally to estimation of any of the submodels considered in the paper. We do not apply the expanded model to the numerical data.

Specifically, the pricing errors are now assumed to be serially correlated across the days and heterogeneous, with variation across contracts being attributed both to 'day of the week' effects and differences arising from the moneyness of particular contracts. Observed option prices C_{ijt} are sorted into moneyness groups according to

$$m_j < \frac{S_{ijt}}{K_{ijt}} \leq m_{j+1},$$

for $j = 1, ..., J$, $m_0 = 0$. Where previously option prices and their associated determinants were only indexed by i, the notation is now expanded to include subscripts indicating contract i in group j on day t. The number of groups and the location of the segment boundaries are chosen so as to accommodate large differences attributable to moneyness. These specifications result in the following model for the observed option prices,

$$C_{ijt} = b_{0j} + b_{1j} q_H(z_{ijt}, \phi) + \sum_{l=1}^{4} d_{lj} D_l + \sum_{g=1}^{G} r_{gj} \overline{C}_{\cdot j(t-g)} + \sigma_j u_{ijt}, \tag{56}$$

where $u_{ijt} \sim N(0,1)$, $i = 1, 2, \ldots, n_{jt}$, $j = 1, 2, ..., J$ and $t = 1, 2, \ldots, n$, with n_{jt} denoting the number of contracts traded in moneyness group j on day t. The D_l, $l = 1, 2, \ldots, 4$, are 'dummy' indicator variables used to represent the day of the week, with Friday corresponding to $D_l = 0$ for all l. The symbol $\overline{C}_{\cdot j(t-g)}$ denotes the average of option prices on day $t - g$ of those contracts traded in moneyness group j, $g = 1, 2, \ldots G$, for a maximum of G lags. When the average price within moneyness group j is not available due to a lack of trades on a particular day, a value is obtained via linear interpolation between the average prices calculated on the closest days.

The coefficients to be estimated for each moneyness group include an intercept, b_{0j}, the coefficient of the option theoretical option price, b_{1j}, day of the week coefficients, d_{lj} for $l = 1, \ldots, 4$, and the lag price coefficients, r_{gj} for $g = 1, 2, \ldots G$. These coefficients may be arranged together by moneyness group, and denoted by $\beta_j = (b_{0j}, b_{1j}, d_{1j}, d_{2j}, d_{3j}, d_{4j}, r_{1j}, \ldots, r_{Gj})'$, for $j = 1, \ldots, J$, with $\beta = (\beta_1, \beta_2, \ldots, \beta_J)'$. Each moneyness group also has an unknown variance, σ_j^2 , to be estimated, and the set of variances may be denoted with $\Sigma = diag\left(\sigma_1^2, \ldots, \sigma_J^2\right)$.

Arranging and stacking the observations into moneyness groups, ordered by day and time of trade, the joint density function for the option prices corresponding to (56) is

$$p(c|z,\phi) = (2\pi)^{-N/2} \prod_{j=1}^{J} \sigma_j^{-N_j} \exp\left(-\frac{1}{2\sigma_j^2}\left[Y_j - X_j\beta_j\right]'\left[Y_j - X_j\beta_j\right]\right), \quad (57)$$

where X_j is an $(N_j \times k)$ matrix containing the relevant regressor variables, $k = 6+G$, Y_j is an $(N_j \times 1)$ vector of observed options prices for moneyness group j, and $N_j = \sum_{t=1}^{n} n_{jt}$. Assuming a joint prior for β and Σ of the form

$$p(\beta, \Sigma) \propto \prod_{j=1}^{J} \sigma_j^{-2}, \quad (58)$$

and imposing *a priori* independence between (β, Σ) and $\delta = \{\phi, \mu\}$, the marginal posterior for δ can be derived as

$$p(\delta|c,s) \propto \prod_{j=1}^{J} \left|X_j'X_j\right|^{-1/2} \widehat{\sigma}_j^{-(N_j-k)} \times p(s,v|\omega^a,\rho,\mu) \times p(\omega^a,\rho,\lambda,\mu),$$

where $(N_j - k)\,\widehat{\sigma}_j^2 = \left[Y_j - X_j\widehat{\beta}_j\right]'\left[Y_j - X_j\widehat{\beta}_j\right]$ and $\widehat{\beta}_j = \left(X_j'X_j\right)^{-1}X_j'Y_j$. Note in particular that X_j' is a function of ϕ, as is $\widehat{\sigma}_j^2$.

To accommodate the additional parameters β and Σ in the MCMC algorithm outlined in Section 3.2, the five conditional posterior distributions outlined therein are replaced by the expanded conditionals, $p(v|\omega^a, \lambda, \rho, \mu, \beta, \Sigma, c, s)$, $p(\omega^a|v, \lambda, \rho, \mu, \beta, \Sigma, c, s)$, $p(\lambda|v, \omega^a, \rho, \mu, \beta, \Sigma, c, s)$, $p(\rho|v, \omega^a, \lambda, \mu, \beta, \Sigma, c, s)$ and $p(\mu|v, \omega^a, \lambda, \rho, \beta, \Sigma, c, s)$. Values can be simulated from these conditionals using a similar hybrid algorithm to that described in Section 3.2. Simulation from the sixth conditional distribution, $p(\beta, \Sigma|v, \omega^a, \lambda, \rho, \mu, c, s)$, occurs via standard Bayesian methods. Specifically, J independent pairs $\left(\beta_j, \sigma_j\right)$ are sampled by first sampling σ_j from an inverted gamma distribution with mean equal to $(\sqrt{(N_j-k)/2}\;\Gamma\left[(N_j-k-1)/2\right]\,/\Gamma\left[(N_j-k)/2\right])\widehat{\sigma}_j$ and variance equal to $\widehat{\sigma}_j^2[(N_j-k)/(N_j-k-2) - (\sqrt{(N_j-k)/2}\Gamma\left[(N_j-k-1)/2\right]/\Gamma\left[(N_j-k)/2\right])^2]$, and then sampling β_j from a normal distribution with mean $\widehat{\beta}_j$and variance $\sigma_j^2(X_j'X_j)^{-1}$, for $j = 1, \ldots, J$.

Note that the aim of categorizing the option price observations into moneyness groups is to distinguish between the explanatory ability of the Heston option price model, $q_H(z_{ijt}, \phi)$, and the features of the data that are known to be a function of the moneyness. Similarly, (56) also accommodates differences across trading days and memory across days in terms of pricing, for a given moneyness group. At present, the variation in option prices attributable to factors such as moneyness and day of the week effects has not been explored in great detail and is a direction for future work. The moneyness groups are chosen for convenience, and need not produce groups of equal numbers of observations.

6 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, and can readily cater for structure in option prices over and above that related directly to the theoretical model itself. The numerical scheme adopted exploits the structure of the square root volatility process, as well as the approximately linearity of the relationship between the theoretical option price and the price of volatility risk. 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. Experimentation with artificially generated data suggest that the method does well at recovering the features of the true volatility process, including the risk premium parameter which has been factored into the option prices.

References

[1] Bakshi, G, Cao, C. and Chen, Z. (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 Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy*, 81, 637-659.

[5] Chao, J.C. and Phillips, P.C.B., (1998), "Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys Prior", *Journal of Econometrics*, 87, 49-86.

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

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

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

[9] Eraker, B., (2001), "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices", Draft Paper, Chicago Business School.

[10] Eraker, B., Johannes, M.J. and Polson, N.G., (2001), "The Impact of Jumps in Returns and Volatility", Draft Paper, University of Chicago, Columbia University.

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

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

[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., Prentice Hall, New Jersey.

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

[16] Jones, C., (2000), "The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets", Draft Paper, University of Rochester.

[17] Kleibergen, F., and Van Dijk, H.K., (1994), "On the Shape of the Likelihood/Posterior in Cointegration Models", Econometric Theory, 10, 514-551.

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

[19] Martin, G.M., (2000), "U.S. Deficit Sustainability: a New Approach Based on Multiple Endogenous Breaks", Journal of Applied Econometrics 15, 83-106.

[20] Martin, G.M., (2001), "Bayesian Analysis of a Fractional Cointegration Model", Econometric Reviews, 20, 217-234.

[21] Martin, G.M., Forbes, C.S. and Martin, V.L., (2001), "Implicit Bayesian Inference Using Option Prices", Draft Paper, Monash University.

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

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

[24] Verdinelli, I. and Wasserman, L., (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,

$$\begin{aligned}
& \frac{1}{2}v(t)S^2(t)\frac{\partial^2 U}{\partial S(t)^2} + \rho\sigma_v v(t)S(t)\frac{\partial^2 U}{\partial S(t)\,\partial v(t)} + \frac{1}{2}\sigma_v^2 v(t)\frac{\partial^2 U}{\partial v(t)^2} \\
& + rS(t)\frac{\partial U}{\partial S(t)} + \left[\kappa^a(\theta^a - v(t)) - \lambda v(t)\right]\frac{\partial U}{\partial v(t)} - rU + \frac{\partial U}{\partial t} \\
= \; & 0 \qquad (59)
\end{aligned}$$

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. Hence it is necessary to choose what part will 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

$$dU - \frac{\partial U}{\partial S(t)}dS(t). \qquad (60)$$

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

$$r(U - \frac{\partial U}{\partial S(t)}S(t))dt. \qquad (61)$$

However, since there is no tradable security relating to the stochastic volatility, the expression for (60) 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 $-\lambda v(t)\frac{\partial U}{\partial v(t)}dt$. 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. This leads to equation (59). 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 $d\varepsilon_1(t)$,

$$d\varepsilon_2(t) = \rho d\varepsilon_1(t) + \sqrt{1-\rho^2}dW, \qquad (62)$$

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 (52). In this case, to obtain the Heston equation we must replace the term

$$\sigma_v \frac{\partial U}{\partial v(t)} \left[\rho(r-\mu)dt + \sqrt{1-\rho^2}\sqrt{v(t)}dW \right] \tag{63}$$

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

Equation (59) can be rewritten in the form

$$\begin{aligned} & \frac{1}{2} v(t) S^2(t) \frac{\partial^2 U}{\partial S(t)^2} + \rho \sigma_v v(t) S(t) \frac{\partial^2 U}{\partial S(t)\, \partial v(t)} + \frac{1}{2} \sigma_v^2 v(t) \frac{\partial^2 U}{\partial v(t)^2} \\ & + r S(t) \frac{\partial U}{\partial S(t)} + \kappa(\theta - v(t)) \frac{\partial U}{\partial v(t)} - rU + \frac{\partial U}{\partial t} \\ = \; & 0 \end{aligned} \tag{64}$$

where

$$\kappa = \kappa^a + \lambda, \;\; \theta = \frac{\kappa^a \theta^a}{\kappa^a + \lambda}. \tag{65}$$

This is the form of (59) which would arise in a hypothetical risk neutral world characterized by mean reversion parameter κ and long-run mean parameter θ. For negative values of λ, the variances in the risk neutral world would revert more slowly to a higher mean. Since $\frac{\partial U}{\partial v(t)}$ is positive, negative risk premiums arise if $\lambda < 0$. Given that the empirical evidence in the literature suggests that $\lambda < 0$, this indicates that for an investment that is correlated with volatility, the market is willing to accept a return below the risk-free rate. This is due to the fact that volatility is negatively correlated with most portfolios.

The Heston option price, denoted by q_H in the text, is the solution of (59) subject to the boundary conditions

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

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

$$dx(t) = [r - \frac{1}{2}v(t)]dt + \sqrt{v(t)}d\varepsilon_1(t) \tag{66}$$

and

$$dv(t) = \kappa[\theta - v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t) \tag{67}$$

respectively. The risk neutral method of solution outlined above yields

$$\begin{aligned} q_H &= e^{-r\tau}\int_K^\infty (S(t) - K)\, p^*(S(t))\, dS(t) \\ &= S(t)P_1 - Ke^{-r\tau}P_2, \end{aligned} \tag{68}$$

where

$$P_1 = \int_K^\infty \frac{e^{-r\tau}S(t)}{S(t)}\, p^*(S(t))\, dS(t) \tag{69}$$

and

$$P_2 = \int_K^\infty p^*(S(t))\, dS(t). \tag{70}$$

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

$$dx(t) = [r + \frac{1}{2}v(t)]dt + \sqrt{v(t)}d\varepsilon_1(t) \tag{71}$$

$$dv(t) = [\kappa\theta - (\kappa - \rho\sigma_v)v(t)]dt + \sigma_v\sqrt{v(t)}d\varepsilon_2(t). \tag{72}$$

Although the above discussion indicates that the pdf p^* corresponding to the bivariate process in (66) and (67) 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, are based on simulated variances with a mean value of 0.04 (or a mean for the volatility quantities of 0.2), a value which is fairly typical of volatilities extracted from empirical data.