

Modeling and Predicting Volatility and its Risk Premium: a Bayesian Non-Gaussian State Space Approach

Gael M. Martin, Catherine S. Forbes and Simone Grose

Department of Econometrics and Business Statistics, Monash University

PRELIMINARY AND INCOMPLETE DRAFT ONLY

June 6, 2009

Abstract

The object of this paper is to model and forecast both objective volatility and its associated risk premium using a non-Gaussian state space approach. Option and spot market information on the unobserved volatility process is captured via non-parametric, ‘model-free’ measures of option-implied and spot price-based volatility, with the two measures used to define a bivariate observation equation in the state space model. The risk premium parameter is specified as a conditionally deterministic dynamic process, driven by past ‘observations’ on the volatility risk premium. The inferential approach adopted is Bayesian, implemented via a Markov chain Monte Carlo (MCMC) algorithm that caters for the non-linearities in the model and for the multi-move sampling of the latent volatilities. The simulation output is used to estimate predictive distributions for objective volatility, the instantaneous risk premium and the conditional risk premium associated with a one month option maturity. Linking the volatility risk premium parameter to the risk aversion parameter in a representative agent model, we also produce forecasts of the relative risk aversion of a representative investor. The methodology is applied both to artificially simulated data and to empirical spot and option price data for the S&P500 index over the 2004 to 2006 period.

1 Introduction

Volatility estimates play a central role in financial applications, with accurate forecasts of future volatility being critical for asset pricing, portfolio management and Value at Risk (VaR) calculations. Along with the information on volatility embedded in historical returns on a financial asset, the prices of options written on the asset also shed light on the option market's assessment of the volatility that is expected to prevail over the remaining life of the options. As such, many forecasting exercises have used both sources of market data to extract information on future volatility, with the relative accuracy of the options- and returns-based forecasts being gauged via a variety of means (e.g. Blair, Poon and Taylor, 2001, Neely, 2003, Martens and Zein, 2004, Pong, Shackleton, Taylor and Xu, 2004, Jiang and Tian, 2005, Koopman, Jungbacker and Hol, 2005, and Martin, Reidy and Wright, 2008).

Crucially, as option pricing occurs under the risk-neutralized measure for the underlying asset price process, any systematic disparity between returns- and option-implied volatility forecasts can be viewed as evidence of the option market having factored in non-zero prices for various risk factors, including volatility risk. A recent literature has evolved in which this disparity has been used - in one way or another- to extract information on the volatility risk premium (e.g. Guo, 1998, Chernov and Ghysels, 2000, Pan, 2002, Jones, 2003, Eraker, 2004, Carr and Wu, 2008, Wu, 2005, Forbes, Martin and Wright, 2007 and Bollerslev Gibson and Zhou, 2008, Bollerslev, Sizova and Tauchen, 2009). However, in none of this work has the primary focus been the extraction of the risk premium for the purpose of improving the accuracy with which objective volatility can be forecast from the dual data source.¹

The primary aim of this paper is to combine option and spot price information with a view to producing accurate forecasts of the objective volatility process of the underlying. A non-Gaussian, non-linear state-space framework is used to model volatility and its associated (time-varying) risk premium as latent state variables. Rather than link market price information to the state variables via complex theoretical option price formulae (e.g. Polson and Stroud, 2003, Eraker, 2004, Forbes, Martin and Wright, 2007, and Johannes, Polson and Stroud, 2008), we use direct non-parametric measures of volatility (see Britten-Jones and Neuberger, 2000, Barndorff-Nielsen and Shephard, 2002, Andersen *et al.*, 2003, and Jiang and Tian, 2005) to define a bivariate observation equation. Theoretical results in Britten-Jones and Neuberger (2000), Barndorff-Nielsen and Shephard (2002) and Bollerslev and Zhou (2002) are exploited in order to link the non-parametric measures to the latent volatility - assumed to follow a Heston (1993) square root process.

¹Johannes, Polson and Stroud (2008) attempt to assess the impact of risk premia on the accuracy of volatility forecasts produced from both option and spot price data, but do not formally incorporate estimation of those risk premia into the analysis.

A secondary aim is to forecast the volatility risk premium itself. Exploiting recent theoretical developments in Carr and Wu (2008), we use the observed volatility measures to extract time series information on the parameter which characterizes the instantaneous risk-related adjustment to the drift in the Heston volatility model. Allied with forecasts of the level of latent volatility itself we produce forecasts of both the instantaneous risk premium and the aggregate risk premium associated with a one month option maturity. Linking the risk premium parameter to the risk aversion of the representative investor, we also produce forecasts of risk aversion via a dynamic model for this quantity.

The inferential approach adopted is Bayesian, with the output of the analysis being the probability distribution of the fixed unknown parameters and the time-varying state variables, conditional on the observed option and spot price data. We use a Markov chain Monte Carlo (MCMC) algorithm, based on a modification of the algorithm of Stroud, Muller and Polson (2003) to cater for the non-linearities in the model and the multi-move sampling of the latent states.

The risk premium parameter which features in the risk-neutral volatility process is allowed, as noted above, to be time varying, but in a manner which does not violate arbitrage assumptions. Specifically, a conditionally deterministic process, driven by past ‘observations’ on the risk premium is used to capture the dynamic behaviour of this component of the model. This approach is computationally simple, with the posterior distribution of the risk premium at any time point - including future time points - able to be estimated from the MCMC draws of the fixed parameters to which the premium is functionally related. The eschewing of any explicit dependence of the risk premium on other observed variables also renders high-frequency real-time forecasting operational.²

The remainder of the paper is organized as follows. Section 2 describes the Heston stochastic volatility model, assumed to underlie both the spot and option price data and interprets the volatility risk premium that forms part of that model. Section 3 outlines the state-space approach that we use to analyse the Heston model, including the time-varying volatility risk premium that we embed within it. The link with investor risk aversion and the recent theoretical developments in Carr and Wu (2008) on volatility swaps, is used to motivate a dynamic specification for the risk premium parameter. A description of the MCMC algorithm used to estimate the latent variables and static parameters and to produce the forecasts is provided in Section 4. A simulation experiment, in which the algorithm is applied to artificially generated data, calibrated to spot and option data on the S&P500

²Bollerslev *et al.* (2008) drive the risk premium by various observed macro-finance variables, in addition to lagged values of observed realized volatility and option-implied volatility. This approach is feasible due to the fact that the authors conduct their empirical analysis using monthly data.

index, is detailed in Section 5. The results of an empirical investigation of intraday spot and option price data for the S&P500 index from November 2004 to May 2006 are reported in Section 6.

2 Interpretation of the Volatility Risk Premium

The objective volatility process is assumed to follow the stochastic model in Heston (1993), whereby the underlying spot price P_t , and stochastic variance V_t , evolve according to the following bivariate process,

$$dP_t = \mu P_t dt + \sqrt{V_t} P_t dW_t^P \quad (1)$$

$$dV_t = \kappa[\theta - V_t]dt + \sigma_v \sqrt{V_t} dW_t^v. \quad (2)$$

The variance V_t is a latent state variable with an innovation governed by a Brownian motion W_t^v , $\theta > 0$ is the actual long-run mean of V_t , to which V_t reverts at rate $\kappa > 0$, while $\sigma_v > 0$ measures the volatility of the variance. Imposing the restriction $\sigma_v^2 \leq 2\kappa\theta$ guarantees that V_t stays in the positive region almost surely (see, for instance, Cox *et al.*, 1985). The pair of Brownian motions (W_t^P, W_t^v) can be correlated with a coefficient ρ to capture the so-called leverage effect.

Based on this particular dynamic model, equilibrium arguments (see Breeden, 1979, Cox *et al.*, 1985a) can be used to produce the following risk-neutral distribution,

$$dP_t = r P_t dt + \sqrt{V_t} P_t dW_t^{*P} \quad (3)$$

$$dV_t = \kappa^*[\theta^* - V_t]dt + \sigma_v \sqrt{V_t} dW_t^{*v}, \quad (4)$$

under which options on the underlying asset are priced, and r denotes the risk-free interest rate (assumed constant). Implicit in the move from (1) and (2) to (3) and (4) are the following transformations in the drift components of the two random processes,

$$\begin{aligned} r &= \mu - (\mu - r) \\ \kappa^*[\theta^* - V_t] &= \kappa[\theta - V_t] - \lambda V_t, \end{aligned}$$

where

$$\kappa^* = \kappa + \lambda; \theta^* = \frac{\kappa\theta}{\kappa^*} \quad (5)$$

and λ is a scalar parameter (viewed as constant for the time being).

The terms $(\mu - r)$ and λV_t represent the premia associated with spot price risk and volatility risk respectively, with the value of λ determining the magnitude (and sign) of the premium factored into option prices for the risk associated with the non-traded state

variable, V_t .³ In empirical applications the estimate of λ is typically negative, which implies slower reversion to a higher mean level than under the objective process; i.e. $\kappa^* < \kappa$ and $\theta^* > \theta$. This implies, in turn, that (call) options are priced higher under the risk-neutral measure, on average, than if they had been priced under the objective measure. That is, a negative value for λ implies that investors are willing to pay a premium for options, as a hedge against possible large movements in the spot price.

The linear (in V_t) form of the volatility risk premium commonly adopted in treatments of the Heston model leads to the convenient relationship between the objective and risk-neutral drift parameters for the volatility process in (5)⁴. As demonstrated by Bollerslev *et al.* (2008), under certain assumptions it also allows an approximate relationship to be derived between the parameter λ and the relative risk aversion parameter in a representative agent model, a relationship which, in turn, links a *negative* value for λ with *positive* aversion to risk. To see this link, note that the equilibrium frameworks of Breeden, 1979 and Cox *et al.* (1985b) lead to factor risk premiums that are equal to the negative of the covariance between changes in the factor and the rate of change in the marginal utility of wealth. For the Heston model, this implies that:

$$\lambda V_t dt = -cov_t\left(\frac{dw_t}{w_t}, dV_t\right), \quad (6)$$

where w_t denotes the marginal utility of wealth for the representative investor⁵. Adopting the canonical power utility function:

$$U_t = e^{-\delta t} \frac{W_t^{1-\gamma}}{1-\gamma}, \quad \gamma > 0,$$

where δ denotes the constant subjective discount rate, it follows that:

$$w_t = e^{-\delta t} W_t^{-\gamma}.$$

Proxying wealth by the value of the market portfolio and assuming that the price process in (1) refers to a market stock index, Ito's lemma yields:

³Note, although we are explicitly modelling the variance and the parameter λ is related to variance risk, we follow the convention in the literature of referring to λ as the volatility risk premium parameter.

⁴In addition to the Heston (1993) article itself, see Guo (1998), Pan (2002), Eraker (2004), Forbes *et al.* (2007) and Bollerslev *et al.* (2008) for treatments of the square root volatility model with a linear volatility risk premium.

⁵As demonstrated in Cochrane (2000), in the continuous time framework w_t itself, rather than the ratio of future to current marginal utilities is used as the stochastic discount factor, or pricing kernel, in asset pricing models.

$$\begin{aligned}
dw_t &= \left[\frac{dw_t}{dP_t} \mu P_t + \frac{dw_t}{dt} + \frac{1}{2} \frac{d^2 w_t}{dP_t^2} V_t P_t^2 \right] dt + \frac{dw_t}{dP_t} \sqrt{V_t} P_t dW_t^p \\
&= \left[-\gamma \frac{e^{-t\delta}}{P_t^{\gamma+1}} \mu P_t - \delta w_t + \frac{1}{2} \frac{1}{P_t^{\gamma+2}} (\gamma e^{-t\delta} + \gamma^2 e^{-2t\delta}) V_t P_t^2 \right] dt \\
&\quad - \gamma \frac{e^{-t\delta}}{P_t^{\gamma+1}} \sqrt{V_t} P_t dW_t^p,
\end{aligned}$$

which gives:

$$\frac{dw_t}{w_t} = \left[-\gamma\mu - \delta + \frac{1}{2}\gamma(1 + \gamma) \right] dt - \gamma\sqrt{V_t} dW_t^p.$$

Using (2) we have:

$$\begin{aligned}
cov_t\left(\frac{dw_t}{w_t}, dV_t\right) &= cov_t(-\gamma\sqrt{V_t}dW_t^p, \sigma_v\sqrt{V_t}dW_t^v) \\
&= -\gamma\sigma_v\rho V_t dt \\
&= -\lambda V_t dt,
\end{aligned}$$

which implies that:

$$\lambda = \gamma\sigma_v\rho. \tag{7}$$

Given $\sigma_v > 0$, $\gamma > 0$ and $\rho < 0$, with the negative value of ρ reflecting the stylized empirical result of volatility increasing as the market declines, $-\lambda$ is a scaled version of the positive risk aversion of the representative agent, as captured by γ . Hence, given estimates of σ_v and ρ , an estimate of λ can be used to produce an estimate of γ .

As is clear from (5), observed option prices, assumed to be priced according to (3) and (4), can be used to identify the parameters of the *objective* process, and the risk premium parameter λ , only if additional information on the objective parameters and/or λ , is introduced. Previous analyses have solved this identification problem: by *jointly* estimating the objective and risk-neutral processes using option and spot price data (e.g. Chernov and Ghysels, 2000, Pan, 2002, Polson and Stroud, 2003, Eraker, 2004, and Forbes, Martin and Wright, 2007); by using option price data only to estimate (3) and (4), and extracting estimates of λ via separate return-based estimates of the objective parameters (Guo, 1998); or by imposing theoretical restrictions on λ (Bates, 2000)⁶. Most importantly, in all of these studies, the link between observed market option prices and the underlying model in (3) and (4) occurs *indirectly*, via a parametric theoretical option price formula derived, in turn, as the expected value of the discounted payoff of the option under the risk-neutral measure.

⁶Note that Bates does not formally estimate the objective parameters. However, the theoretical bounds placed λ would allow bounds to be placed on the values of the objective parameters, given estimates of the risk neutral parameters.

In contrast, we link the observed option price data to the model in (3) and (4) *directly*, by using a non-parametric estimate of expected integrated volatility over the life of the option, evaluated according to the risk-neutral process in (4). Analogously, the observed spot data is linked to the objective process in (1) and (2) by using realized volatility to estimate the integrated volatility associated with the objective process.

In Section 3.1 we outline the state space model based on the volatility measures, with the risk premium parameter λ and, by the above arguments, risk aversion, assumed to be constant. In Section 3.2 we extend the model to allow for a dynamic model for λ . In common parlance, the model we adopt is *observation-driven*, with the value of λ at time t , λ_t , given by a deterministic function of λ_{t-1} and the ‘observed’ value of λ_{t-1} , denoted by l_{t-1} . In specifying l_{t-1} , we exploit recent theoretical developments in Carr and Wu (2008), in which the difference between the two observed measures of volatility is linked to the risk premium parameter via a particular non-linear function.

3 A State Space Model Based on Realized Volatility and Model-free Implied Volatility Measurements

3.1 Constant risk aversion

Given the objective volatility process in (2), we define integrated volatility over the horizon $t - 1$ to t (call this day t) as

$$\mathcal{V}_{t-1,t} = \int_{t-1}^t V_s ds. \quad (8)$$

Denoting by p_{t_i} the i th logarithmic price observed during day t , and $r_{t_i} = p_{t_i} - p_{t_{i-1}}$ as the i th transaction return, it is now standard knowledge (Barndorff-Nielsen and Shephard, 2002, and Anderson et al., 2003) that

$$RV_t = \sum_{t_i \in [t-1,t]}^N r_{t_i}^2 \xrightarrow{p} \mathcal{V}_{t-1,t}, \quad (9)$$

where RV_t is referred as realized volatility and N is equal to the number of intraday returns on day t ⁷. Integrated volatility, $\mathcal{V}_{t-1,t}$, is thus the volatility that will prevail over the horizon of one day, according to (2), and RV_t is a consistent estimator of that volatility. Note that the random nature of V_s implies that $\mathcal{V}_{t-1,t}$ is also random.

⁷As is common, we use the term ‘volatility’ to refer to either a variance or a standard deviation quantity. Which type of quantity is being referenced in any particular instance will be made clear by both the context and the notation. Implicit in the result in (9) is the assumption that microstructure noise effects are absent. The formal incorporation of microstructure noise in the assumed process for intraday returns has led to modifications of RV_t which are consistent estimators of $\mathcal{V}_{t-1,t}$ in the presence of such noise; see Martin *et al.* (2009) for a recent summary.

We propose the measurement equation

$$RV_t = \mathcal{V}_{t-1,t} + u_{RV_t}, \quad (10)$$

where the latent volatility evolves according to (2) and $u_{RV_t} = RV_t - \mathcal{V}_{t-1,t}$ is the realized volatility error. Based on the limit theory in Barndorff-Nielsen and Shephard (2002), we approximate u_{RV_t} as

$$u_{RV_t} \sim N\left[0, \frac{2 \int_{t-1}^t V_s^2 ds}{N}\right]. \quad (11)$$

Hence, we assume a (conditionally) Gaussian measurement equation that is non-linear in the point-in-time state volatility. The state equation also has a non-linear form (in the past states) due to the square root feature of (2).⁸

As well as having spot-price based observations on the latent variance, we have option-based measurements via the following logic. Bollerslev and Zhou (2002) (amongst others) derive a set of conditional moments for the integrated volatility of Heston's (1993) stochastic volatility model. Defining $\mathcal{F}_t = \sigma\{V_s; s \leq t\}$ as the sigma-algebra generated by the point-in-time volatility process, the conditional mean for integrated volatility under the physical measure (2) can be expressed as a linear function of the point-in-time volatility,

$$E(\mathcal{V}_{t,t+\tau} | \mathcal{F}_t) = E\left(\int_t^{t+\tau} V_s ds \middle| \mathcal{F}_t\right) = a_\tau V_t + b_\tau, \quad (12)$$

where

$$a_\tau = \frac{1}{\kappa} (1 - e^{-\tau\kappa}) \quad \text{and} \quad b_\tau = \tau\theta - \frac{\theta}{\kappa} (1 - e^{-\tau\kappa}). \quad (13)$$

Correspondingly, under the distribution in (4), a risk-neutral expectation of integrated volatility over the horizon t to $(t + \tau)$ is given by

$$E^*(\mathcal{V}_{t,t+\tau} | \mathcal{F}_t) = a_\tau^* V_t + b_\tau^*, \quad (14)$$

where

$$a_\tau^* = \frac{1}{\kappa^*} (1 - e^{-\tau\kappa^*}) \quad \text{and} \quad b_\tau^* = \tau\theta^* - \frac{\theta^*}{\kappa^*} (1 - e^{-\tau\kappa^*}). \quad (15)$$

and θ^* and κ^* are as defined in (5).

⁸One of the aims of Barndorff-Nielsen and Shephard (2002) is to see if the prediction of $\mathcal{V}_{t-1,t}$ can be improved upon by using a parametric model for V_s in addition to the measurement RV_t , rather than just using the "raw" RV_t . The Kalman filter is used to define the likelihood function and quasi-maximum likelihood (QML) applied to estimate the unknown fixed parameters. Creal (2007) extends the analysis of Barndorff-Nielsen and Shephard to investigate the improvements that can be had by using particle filtering, which exploits the non-Gaussian nature of the volatility state equation. Neither paper uses the precise form of (11) in the specification of the measurement equation, nor gives any attention to the role of option-implied estimation of the volatility process.

As shown by Britten-Jones and Neuberger (2000), Jiang and Tian (2005) and Carr and Wu (2008), the risk-neutral expectation in (14) is also implied by a continuum (over strike K) of option prices with maturity $\tau > 0$, as

$$E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) = 2 \int_0^\infty \frac{C(t+\tau, K) - C(t, K)}{K^2} dK. \quad (16)$$

This equality is shown to hold for general diffusion processes (extended to jump diffusion processes in Jiang and Tian, 2005), including the type specified in (3) and (4). Hence, $E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$ is referred as “model free” implied volatility. Importantly, this measure eschews the dependence of the ubiquitous Black-Scholes option-implied volatility on the empirically invalid assumption of geometric Brownian motion for the underlying asset price.

Given an estimate of $E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$ in (16), as based on a finite set of observed option prices on day t , which we denote by $MF_{t,t+\tau}$, we can define the following option-based measurement equation,

$$MF_{t,t+\tau} = E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) + u_{MF_t}, \quad (17)$$

where u_{MF_t} captures the error associated with the discretization and truncation of the integral in (16). We assume, as a first pass, that u_{MF_t} is a Gaussian white noise process, and independent of $E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$. Hence, we have a second measurement equation that is linear in V_t and with an additive Gaussian error that is independent of the state process.

Using an Euler discretization of (2), we have

$$RV_t = \mathcal{V}_{t-1,t} + \sqrt{f(V_t)}\xi_{1t} \quad (18)$$

$$\begin{aligned} MF_{t,t+\tau} &= E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) + \sigma_{MF}\xi_{2t} \\ &= \tau\theta^* - \frac{\theta^*}{\kappa^*} (1 - e^{-\tau\kappa^*}) + \frac{1}{\kappa^*} (1 - e^{-\tau\kappa^*}) V_t + \sigma_{MF}\xi_{2t}, \end{aligned} \quad (19)$$

and

$$V_{t+1} = \kappa\theta\Delta t + (1 - \kappa\Delta t) V_t + \sigma_v\sqrt{\Delta t}\sqrt{V_t}\xi_{3t}, \quad (20)$$

with

$$f(V_t) = \frac{2 \int_{t-1}^t V_s^2 ds}{N} \quad (21)$$

and

$$\xi_t = (\xi_{1t}, \xi_{2t}, \xi_{3t})' \stackrel{iid}{\sim} N(0_3, I_3) \text{ for all } t = 1, 2, \dots, T. \quad (22)$$

Setting $\Delta t = 1$, the state equation for V_t describes the evolution of the point-in-time (annualized) volatility from one day to the next. It is this volatility quantity at time t that enters the function $E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$ in (19). In (18), on the other hand, we have $\mathcal{V}_{t-1,t} =$

$\int_{t-1}^t V_s ds$. Again, using the daily interval, a crude approximation to the integrated volatility associated with day t is V_t . We can, however, set $\Delta t = 1/B$ in (20), for $B > 1$. In this case, we have $V_{t-1+i\Delta t}$ = the point-in-time volatility on day $(t-1)$ after i intraday recursions of (20), and $\mathcal{V}_{t-1,t}$ is approximated by $\frac{1}{B} \sum_{i=1}^B V_{t-1+i\Delta t}$, and the integrated quarticity, $\int_{t-1}^t V_s^2 ds$, approximated by $\frac{1}{B} \sum_{i=1}^B V_{t-1+i\Delta t}^2$. Note that θ remains an annualized quantity at all times, matching the annualized magnitude of the point in time volatility, V_s . The parameter κ is treated as a daily quantity, measuring the rate of mean reversion in the annualized V_s per day. In accordance with this treatment of κ , $\tau = 22$ days and $MF_{t,t+\tau}$ is modelled as an aggregated annualized variance over the trading month.

3.2 A dynamic model for risk aversion

Carr and Wu (2008) propose a method of quantifying the volatility risk premium using variance swaps⁹. A variance swap is an over-the-counter contract with a payoff equal to the difference between quadratic variation, defined over the life of the swap contract, and the so-called variance swap rate, which is determined at time t . In the context of the Heston stochastic volatility model, in which jumps in the asset price are not modelled, quadratic variation reduces to integrated volatility, $\mathcal{V}_{t,t+\tau}$, defined in accordance with (8).

Defining $t+\tau$ to be the period at which the contract expires and denoting the price of the variance swap as p_t and its payoff as $x_{t+\tau}$, the no arbitrage conditions that underlie standard asset pricing theory imply that:

$$p_t = cE_t^*(x_{t+\tau}),$$

where E_t^* denotes a conditional risk-neutral expectation, $E_t^* = E^*(\cdot|\mathcal{F}_t)$, and c is a constant (risk-neutral) discount factor. Defining the variance swap rate as $SW_{t,t+\tau}$ we have

$$x_{t+\tau} = \mathcal{V}_{t,t+\tau} - SW_{t,t+\tau}.$$

Given that the variance swap has zero market value at time t , it follows that $p_t = 0$ and

$$SW_{t,t+\tau} = E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) \tag{23}$$

as a consequence. As is consistent with the result in (16), Carr and Wu (2008) show that $SW_{t,t+\tau}$ can indeed be synthesized by a linear combination of τ -maturity option prices observed on day t .

⁹Once again we point out that, despite our terminology, we are actually concerned with the *variance* risk premium here; hence the relevance of *variance* swaps. In an earlier version of their paper Carr and Wu (2004) make the explicit distinction between using variance and volatility swaps, depending on whether the focus is on the variance or volatility risk premium.

In addition to the equality in (23), asset pricing theory allows the zero price of the variance swap to be linked to $x_{t+\tau} = \mathcal{V}_{t,t+\tau} - SW_{t,t+\tau}$ via the objective measure as

$$0 = E(m_{t,t+\tau}x_{t+\tau}|\mathcal{F}_t), \quad (24)$$

where $E(\cdot|\mathcal{F}_t)$ denotes a conditional expectation with respect to the objective measure and

$$m_{t,t+\tau} = \frac{M_{t,t+\tau}}{E_t(M_{t,t+\tau})}$$

is the normalized stochastic discount factor (or pricing kernel), with

$$E_t(M_{t,t+\tau}) = e^{-r\tau}$$

under the assumption of a constant risk-free interest rate. Given that $SW_{t,t+\tau}$ is known at time t , and using $E(m_{t,t+\tau}|\mathcal{F}_t) = 1$, (24) can be re-written as

$$\begin{aligned} SW_{t,t+\tau} &= E(m_{t,t+\tau}\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) \\ &= E(m_{t,t+\tau}|\mathcal{F}_t)E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) + cov(m_{t,t+\tau}\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) \\ &= E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) + cov(m_{t,t+\tau}\mathcal{V}_{t,t+\tau}|\mathcal{F}_t). \end{aligned}$$

Dividing through by $SW_{t,t+\tau}$, we produce an expression for the expected excess return on the variance swap investment

$$E\left(\frac{\mathcal{V}_{t,t+\tau}}{SW_{t,t+\tau}}|\mathcal{F}_t\right) - 1 = -cov\left(m_{t,t+\tau}\frac{\mathcal{V}_{t,t+\tau}}{SW_{t,t+\tau}}|\mathcal{F}_t\right).$$

Hence, an average of

$$\frac{\mathcal{V}_{t,t+\tau}}{SW_{t,t+\tau}} - 1$$

over t represents an estimate of the (return) premium for volatility risk associated with the contract life τ . Alternatively, we can define the premium in variance payoff units as

$$E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - SW_{t,t+\tau} = -cov(m_{t,t+\tau}\mathcal{V}_{t,t+\tau}|\mathcal{F}_t),$$

and estimate the premium via an average of

$$\mathcal{V}_{t,t+\tau} - SW_{t,t+\tau}.$$

Importantly, the above analysis highlights the fact that the *conditional* volatility risk premium defined over the maturity period τ is given by the following linear function of the point in time latent variance V_t ,

$$\begin{aligned} E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - SW_{t,t+\tau} &= E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) \\ &= a_\tau V_t + b_\tau - [a_\tau^* V_t + b_\tau^*], \end{aligned} \quad (25)$$

with the terms on the R.H.S. of (25) defined in (13) and (15). For given sample estimates of $E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$ and $SW_{t,t+\tau}$, and for given values of V_t and the objective parameters, κ and θ , the equation in (25) can be solved for λ , the parameter associated with the underlying (instantaneous) volatility risk premium, $\lambda V_t dt$.¹⁰ In particular, a time series of the difference between sample estimates of $E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$ and $SW_{t,t+\tau}$ can be used to shed light on the dynamic behaviour of λ . To this end, we use a time series plot, over the particular 2 year period associated with the empirical exercise in Section 6, of the values of λ so defined as a preliminary diagnostic of the dynamic properties of the risk premium parameter. As an estimate of the objective conditional expectation, $E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$, $\widehat{E}(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)$, we use the aggregate of the j -step-ahead forecasts, $j = 1, 2, \dots, 22$, produced from an autoregressive fractionally integrated moving average (ARFIMA) model of order $(1,d,1)$ fitted to RV_t using rolling samples of size 1000 up to time t . The estimate of $SW_{t,t+\tau}$ is $MF_{t,t+\tau}$, as based on a finite number of τ -maturity option prices on day t , and as calculated in the manner detailed in Section 6. We denote by l_t the solution for λ_t of the non-linear equation

$$\widehat{E}(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - MF_{t,t+\tau} = f(\lambda_t) \quad (26)$$

at each point t , where l_t represents an ‘observed’ value of λ_t , based on the replacement of the conditional (unobserved) risk premium, $E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - SW_{t,t+\tau}$, with the estimated risk premium, $\widehat{E}(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - MF_{t,t+\tau}$. The function $f(\cdot)$ is given by

$$f(\lambda_t) = a_\tau V_t + b_\tau - [a_{t,\tau}^* V_t + b_{t,\tau}^*], \quad (27)$$

with

$$a_{t,\tau}^* = \frac{1}{\kappa_t^*} (1 - e^{-\tau \kappa_t^*}) \quad \text{and} \quad b_{t,\tau}^* = \tau \theta_t^* - \frac{\theta_t^*}{\kappa_t^*} (1 - e^{-\tau \kappa_t^*}) \quad (28)$$

and

$$\kappa_t^* = \kappa + \lambda_t; \quad \theta_t^* = \frac{\theta \kappa}{\kappa_t^*}. \quad (29)$$

¹⁰Note that the *unconditional* risk premium is given by:

$$\begin{aligned} & E[E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - SW_{t,t+\tau}] \\ &= E[E(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t) - E^*(\mathcal{V}_{t,t+\tau}|\mathcal{F}_t)] \\ &= E[a_\tau V_t + b_\tau - [a_\tau^* V_t + b_\tau^*]] \\ &= E[b_\tau - b_\tau^* + (a_\tau - a_\tau^*) V_t] \\ &= b_\tau - b_\tau^* + (a_\tau - a_\tau^*) \theta \\ &= \frac{\lambda \theta}{\kappa^*} \left[\tau - \frac{1}{\kappa^*} (1 - e^{-\tau \kappa^*}) \right]. \end{aligned}$$

Hence, the unconditional mean of the aggregate risk premium (over τ) defined in this way is negative if λ is negative. This result corresponds correctly to the spot price-based measure, $\mathcal{V}_{t,t+\tau}$, being less, on average, than the option-price-based measure, $SW_{t,t+\tau}$.

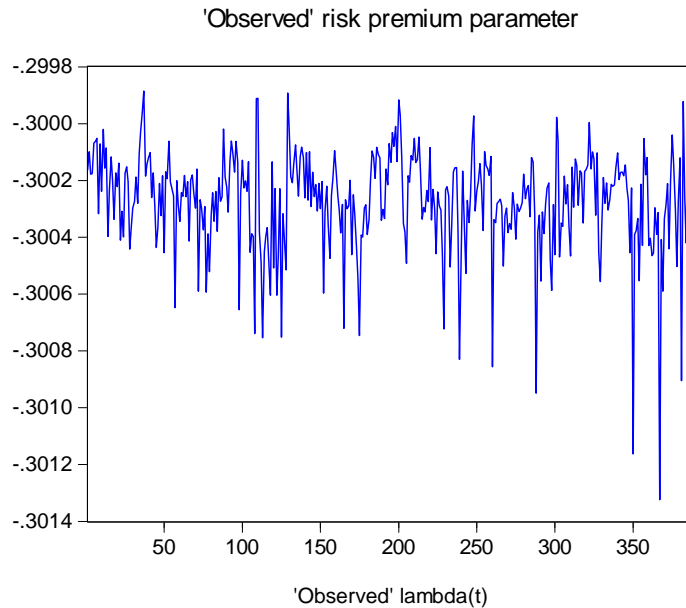


Figure 1: ‘Observed’ Volatility Risk Premium Parameter: l_t

The values inserted into $f(\cdot)$ for κ and θ are based on estimates produced by a preliminary run of the MCMC algorithm using the empirical S&P500 data, with RV_t used as a proxy for the latent V_t .

As is clear from Figure 1, over this particular period of time there is some time series dependence in l_t , evident despite the noise introduced into the calculation by using RV_t as a proxy for V_t . This is confirmed by the sample autocorrelation function for l_t , displayed in Figure 2, which exhibits some significant autocorrelation at the lower lags that is consistent with a short memory process¹¹.

In order to capture dynamic behaviour in the risk premium parameter, we specify a conditionally deterministic specification which mimics a generalized autoregressive heteroscedastic (GARCH) structure for volatility, namely,

$$\lambda_t = \alpha_0 + \alpha_1 \lambda_{t-1} + \delta l_{t-1}, \quad (30)$$

where l_{t-1} denotes the observed value of λ_t at time $t - 1$.¹² This specification for λ_t also mimics the structure of the autoregressive conditional duration (ACD) model for durations (Engle and Russell, 1998) and the observation-driven model for count data analysed in

¹¹See Bollerslev et al. (2009) for related discussion on the dynamics of quantities related to the volatility risk premium.

¹²We assume that $\lambda_0 = \alpha_0 / (1 - (\alpha_1 + \delta_1))$.

Sample autocorrelation function for 'observed' risk premium parameter

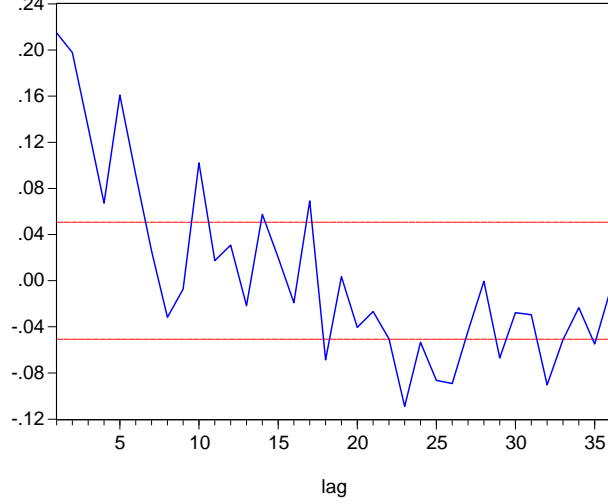


Figure 2: Sample ACF of l_t

Heinen (2003), Jung *et al.* (2006) and Feigen *et al.* (2008). The specification has the advantage of not introducing an additional random risk factor into the model to be priced. It is also advantageous from an inferential point of view, being a function of a small number of fixed parameters only.¹³ With the model extended to allow for a time varying risk premium parameter, the full state space specification becomes (with some repetition here for convenience)

$$RV_t = \mathcal{V}_{t-1,t} + \sqrt{f(V_t)}\xi_{1t}; \quad f(V_t) = \frac{2 \int_{t-1}^t V_s^2 ds}{N} \quad (31)$$

$$MF_{t,t+\tau} = \tau\theta_t^* - \frac{\theta_t^*}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) + \frac{1}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) V_t + \sigma_{MF}\xi_{2t} \quad (32)$$

$$V_{t+1} = \kappa\theta\Delta t + (1 - \kappa\Delta t) V_t + \sigma_v\sqrt{\Delta t}\sqrt{V_t}\xi_{3t} \quad (33)$$

$$\lambda_t = \alpha_0 + \alpha_1\lambda_{t-1} + \delta l_{t-1}, \quad (34)$$

with θ_t^* and κ_t^* as defined in (29) and the distributional assumptions for the innovations as defined in (22). The quantity l_t is an approximate solution of (26). Details of that solution are provided in the following section.

¹³Extensions of this particular model to various other forms of time series models, in particular those that allow for regime shifts and/or threshold effects, will be explored in due course.

In modelling the time-varying parameter λ_t we will, via (7), be effectively modelling time variation in the risk aversion of the representative investor, γ ; see Bollerslev *et al.* (2008), and other references cited therein, for alternative approaches to allowing for non-constant risk-aversion. The output of our Bayesian analysis will include an estimate of the predictive distributions of future values of γ which, as well as being of interest in their own right, can also be used as input into other models in which predictions of risk-aversion are required; e.g. Bollerslev *et al.* (2008). Moreover, given MCMC draws from the predictive distributions of both λ_{T+1} and V_{T+1} , in addition to draws of the fixed parameters, we can, via (25), produce draws from the predictive distribution of the conditional volatility risk premium. Predictive accuracy can then be assessed by applying various scoring rules (see, for example, Gneiting *et al.* 2007) evaluated at the *observed* deviation in realized and option implied volatility over the option maturity period, namely, $RV_{T+1,T+1+\tau} - MF_{T+1,T+1+\tau}$.

4 MCMC Algorithm

Given the complexity of the state space model represented by (31) to (34), the joint posterior distribution for all unknowns is analytically intractable. Hence, an MCMC algorithm may be applied to produce draws from the joint posterior and those draws then used to estimate inferential quantities of interest, including predictive densities, in the usual way. The algorithm described here is based on the simplifying assumption of a constant variance for the measurement error in (31), i.e. that $f(V_t) = \sigma_{RV}^2$. A state dependent variance will be accommodated in due course. The approximation in of $\mathcal{V}_{t-1,t}$ in (31) by V_t is also invoked for the time being. The joint posterior density for all unknowns is denoted by $p(V, \phi | MF, RV)$, where the vector of latent volatilities is given by $V = (V_1, V_2, \dots, V_T)'$ and the vector of fixed parameters by $\phi = (\kappa, \theta, \sigma_{MF}, \sigma_{RV}, \sigma_v, \alpha_0, \alpha_1, \delta)'$. The observed data vectors are denoted by $RV = (RV_1, RV_2, \dots, RV_T)'$ and $MF = (MF_{1,1+\tau}, MF_{2,2+\tau}, \dots, MF_{T,T+\tau})'$. The posterior density satisfies

$$\begin{aligned}
p(V, \phi | MF, RV) &\propto \left[\prod_{t=2}^T p(MF_{t,t+\tau} | MF_{t-1,t-1+\tau}, RV_{t-1}, RV_{t-2}, \dots, RV_{t-\tau}, V_t, \kappa, \theta, \sigma_{MF}, \alpha_0, \alpha_1, \delta) \right. \\
&\quad \times p(RV_t | V_t, \sigma_{RV}) \times p(V_t | V_{t-1}, \kappa, \theta, \sigma_v) \left. \right] \\
&\quad \times p(\kappa, \theta, \sigma_{MF}, \sigma_{RV}, \sigma_v, \alpha_0, \alpha_1, \delta), \tag{35}
\end{aligned}$$

where it is assumed that $V_1 = \theta$.

The conditioning of $MF_{t,t+\tau}$ on $MF_{t-1,t-1+\tau}$ and $RV_{t-1}, RV_{t-2}, \dots, RV_{t-\tau}$ in the first component on the R.H.S. of (35) derives from the dependence of the measure $MF_{t,t+\tau}$ on l_{t-1} , via the model for λ_t in (34). The value of l_t is, in turn, produced at each t , by taking a

first-order Taylor series expansion of $f(\cdot)$ in (27),

$$f(l_t) = f(l) + f'(l_t)|_{l_t=l} (l_t - l)$$

around a fixed value l (to be specified below) and solving for l_t as

$$l_t = \frac{RV_{t-\tau,t} - MF_{t,t+\tau} - [a_\tau V_t + b_\tau] + [a_\tau^*(l)V_t + b_\tau^*(l)]}{f'(l_t)|_{l_t=l}} + l, \quad (36)$$

where

$$f'(l_t)|_{l_t=l} = \frac{1}{[\kappa^*(l)]^2} (1 - e^{-\tau\kappa^*(l)}) (2\theta^*(l) - V_t) + \frac{1}{\kappa^*(l)} \tau e^{-\tau\kappa^*(l)} (V_t - \theta^*(l)) - \frac{\tau\theta^*(l)}{\kappa^*(l)},$$

with $\kappa^*(l)$ and $\theta^*(l)$ given by the expressions in (5), based on $\lambda = l$, and $a_\tau^*(l)$ and $b_\tau^*(l)$ are defined as in (15), with $\kappa^* = \kappa^*(l)$ and $\theta^* = \theta^*(l)$.

The Gibbs-based MCMC algorithm is implemented in two main steps:

1. Generating V from

$$p(V|\phi, MF, RV) \propto \prod_{t=2}^T p(MF_{t,t+\tau} | MF_{t-1,t-1+\tau}, RV_{t-1}, RV_{t-2}, \dots, RV_{t-\tau}, V_t, \kappa, \theta, \sigma_{MF}, \alpha_0, \alpha_1, \delta) \\ \times p(V_t | V_{t-1}, \kappa, \theta, \sigma_v).$$

2. Generating ϕ (with elements of ϕ blocked conveniently) from

$$p(\phi|V, MF, RV) \propto \left[\prod_{t=2}^T p(MF_{t,t+\tau} | MF_{t-1,t-1+\tau}, RV_{t-1}, RV_{t-2}, \dots, RV_{t-\tau}, V_t, \kappa, \theta, \sigma_{MF}, \alpha_0, \alpha_1, \delta) \right] \\ \times p(RV_t | V_t, \sigma_{RV}) p(V_t | V_{t-1}, \kappa, \theta, \sigma_v) \times p(\kappa, \theta, \sigma_{MF}, \sigma_{RV}, \sigma_v, \alpha_0, \alpha_1, \delta).$$

With standard non-informative priors being invoked for all parameters, the simulation of the individual standard deviation parameters, σ_{MF} , σ_{RV} and σ_v , is standard, via inverted gamma distributions respectively. The conditional posteriors of κ , θ and the parameters of the process for $\lambda_t - \alpha_0$, α_1 and δ - on the other hand, are non-standard due to the fact that the conditional mean function in (19) is non-linear in all parameters. In the case of both κ and θ we use the structure of the model to define Gaussian kernels used, in turn, to produce candidate draws for MH sub-steps. In the case of α_0 , α_1 and δ we insert random walk MH sub-steps. Once draws of all fixed parameters have been obtained, and based on a starting value, $\lambda_0 = \alpha_0 / (1 - [\alpha_1 + \delta])$, draws of the vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_T)'$ are produced automatically from (the degenerate) $p(\lambda|V, \phi, MF, RV)$ via the conditionally deterministic relationship in (34). Details of the mixture-based Metropolis Hastings (MH) algorithm used to draw V (with random blocking) are provided in Appendix A. Algorithmic details related to ϕ are given in Appendix B.¹⁴

¹⁴The numerical results reported in the following sections have been produced (in the main) using the

5 Simulation Experiment

In order to determine the accuracy with which the Bayesian method estimates the model parameters, the stochastic variances and the time-varying risk premium, we conduct a small Monte Carlo experiment based on parameter values calibrated (approximately) to empirical S&P500 data (of the type analysed in Section 6) over the mid-2001 to mid-2006 period. This period was chosen as one in which both volatility measures were (in the earlier sub-period at least) very high and very persistent, corresponding to values of θ and κ , respectively, rather high and low. We first simulate V_s from (33) over 30 second intervals. Assuming a trading day that matches that underlying the empirical data, namely from 10.00am to 4.00pm, a period of 30 seconds translates into $\Delta t = 30/(6 \times 60 \times 60) = 0.0014$. Using a corresponding Euler approximation of (1), the 720 values of V_s are then used to produce intraday prices, which are then transformed into intraday returns. The sum of squared returns over the 30 second intervals is then used as a proxy for the true unobservable integrated volatility on the right hand side of (31). Generation of the T values of $MF_{t,t+\tau}$ occurs via (32), with $\tau = 22$. The risk premium is assumed to be dynamic as per (34).

Based on the artificially generated data, the Bayesian algorithm is applied assuming $\Delta t = 1$; that is, a daily discretization is used, with $\int_{t-1}^t V_s^2 ds$ in (31) approximated by V_t accordingly. The sample size of the simulation experiment is set equal to $T = 2000$. Table 1 reports the value of each parameter used in producing the simulated data along with the relevant estimated marginal posterior mean and 95% highest posterior density (HPD) interval, based on $B = 5000$ iterations after a burn-in of 5000. In addition, selected values of the latent stochastic variance and λ_t values are also reported, along with the corresponding marginal posterior mean and 95% HPD intervals. The estimation results from this simulation confirm that the estimation methodology works well, at least for the selected parameter settings. In the bottom panel of Table 1 we report the true (simulated) future values of the latent variables, V_{T+1} and λ_{T+1} , along with the predictive means and the 95% prediction intervals.

In addition to the desired forecast quantities, time series plots of the simulated V_t series and posterior mean of the daily V_t quantities estimated via the MCMC algorithm are shown in Figure 3. It is apparent from this graph that the dynamic pattern in the latent daily volatility series is captured by the posterior mean. Over the $T = 2000$ observations, the sample correlation between the ‘true’ V_t process and its estimated posterior mean value is 95.7%.

JAVA programming language. The authors would like to acknowledge the contribution of Alex Nichols to the writing of some earlier versions of the programs on which the current results are based.

Table 1: Marginal Posterior Density Results for the Heston Model: Artificially Generated Data with a Dynamic Risk Premium Parameter. Marginal Posterior Estimates Based on 5000 Draws Following a 5000 Iteration Burn-in Period.

Parameter	True Value	Marginal Posterior Mean	95% <i>HPD</i> interval
κ	0.070	0.0708	(0.0623, 0.0801)
θ	0.042	0.0420	(0.0392, 0.0447)
σ_v	0.030	0.0287	(0.0264, 0.0311)
σ_{RV}	0.010	0.0099	(0.0095, 0.0103)
σ_{MF}	0.010	0.0104	(0.0100, 0.0108)
$\alpha_0 (1 - [\alpha_1 - \delta])$	-0.06	-0.0600	(-0.0644, -0.0560)
$\alpha_1 + \delta$	0.80	0.6002	(0.0273, 0.9497)
Variance (V_t)			
V_{100}	0.051489	0.0499	(0.0412, 0.0586)
V_{300}	0.040757	0.0422	(0.0335, 0.0503)
V_{2000}	0.018287	0.0225	(0.0124, 0.0310)
Risk Premium (λ_t)			
λ_{100}	-0.060014	-0.0600	(-0.0652, -0.0564)
λ_{300}	-0.060138	-0.0602	(-0.0659, -0.0565)
λ_{2000}	-0.059954	-0.0598	(-0.0649, -0.0562)
Forecast Values			
V_{2001}	0.015846	0.0237	(0.0117, 0.0359)
λ_{2001}	-0.059889	-0.0598	(-0.0649, -0.0560)

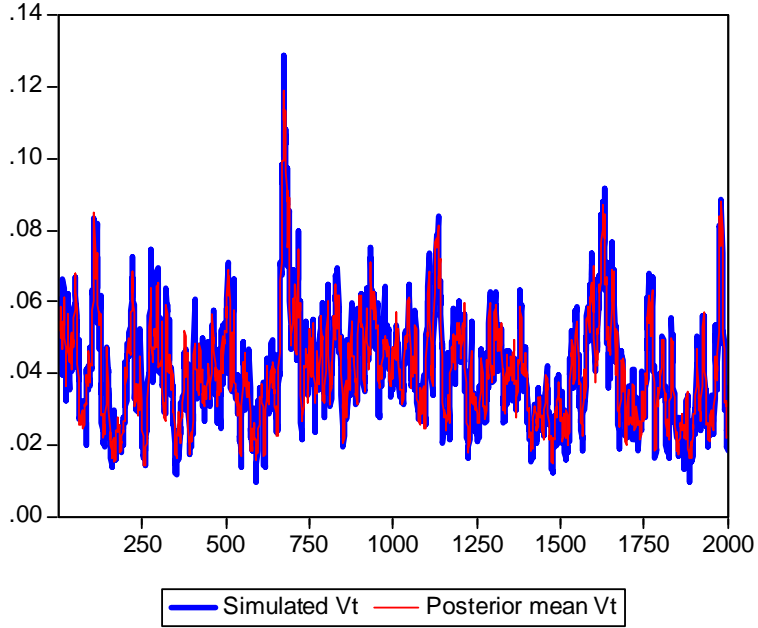


Figure 3: Simulated V_t Process and Marginal Posterior Mean of V_t Process.

As reflected in the summary results in Table 1 for λ_{100} , λ_{300} and λ_{2000} , the posterior means of λ_t , for all $t = 1, 2, \dots, T$, are quite accurate estimates of the true values. However, the results in Table 1 also indicate that the degree of persistence in λ_t (as approximated by $\alpha_1 + \delta$) is slightly underestimated. In order to visualize the relative dynamics in the true and estimated λ_t processes we plot in Figure 4 the true λ_t against the posterior mean at each t , in turn estimated using the Gibbs draws, and also ‘demeaned’ using the time series average of posterior mean estimates. The posterior mean series broadly reflects the fluctuations in the latent λ_t process, but is noisier than the true λ_t series. Over the $T = 2000$ observations, the sample correlation between the ‘true’ λ_t series and its estimated posterior mean value is 92.9%.

Bayesian predictions are produced by estimating the predictive densities:

$$p(V_{T+1}|MF, RV) = \int p(V_{T+1}|V_T, \phi)p(V_T, \phi|MF, RV)dV_Td\phi \quad (37)$$

and

$$p(\lambda_{T+1}|MF, RV) = \int p(\lambda_{T+1}|\lambda_T, \phi, MF, RV)p(\lambda_T, \phi|MF, RV)d\phi, \quad (38)$$

where $p(V_{T+1}|V_T, \phi)$ is a normal pdf with mean and variance defined according to (20), and $p(\lambda_{T+1}|\lambda_T, \phi, MF, RV)$ is, conditional on MF and RV , a degenerate distribution associated

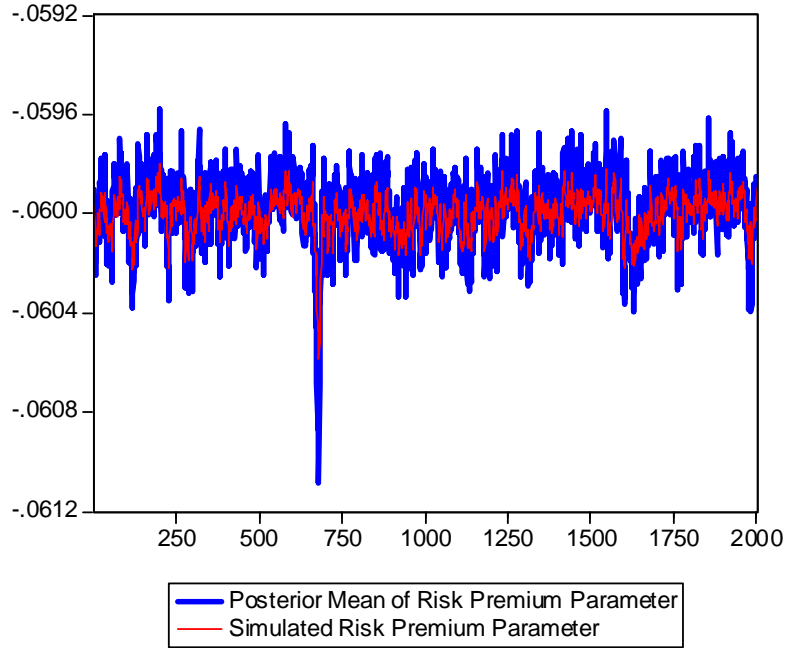


Figure 4: Simulated λ_t Process and Marginal Posterior Mean of λ_t Process

with the deterministic relationship in (30), with l_T equated with the solution of (26) at time T . The predictive density in (37) is estimated by drawing V_{T+1} from

$$p(V_{T+1}|V_T^{(i)}, \phi^{(i)}),$$

for $i = 1, 2, \dots, B$, and applying kernel smoothing to the B draws. This estimated predictive density for V_{T+1} resulting from the simulation experiment is shown in Figure 5. The ‘true’ simulated V_{T+1} value in this case is $V_{2001} = 0.01585$, a value that is well within the 95% HPD interval of the predictive distribution. The same approach is used to estimate the predictive density of λ_{T+1} in (38), with the estimated predictive density for λ_{T+1} resulting from the simulation experiment is shown in Figure 6. Note the conditional density $p(\lambda_{T+1}|\lambda_T, \phi, MF, RV)$ in (38) is degenerate; however integration over $p(\lambda_T, \phi|MF, RV)$ results in a non-degenerate predictive posterior density for λ_{T+1} . The ‘true’ simulated λ_{T+1} value in this case is $\lambda_{2001} = -0.05989$, again a value that is well within the central mass of the predictive distribution.

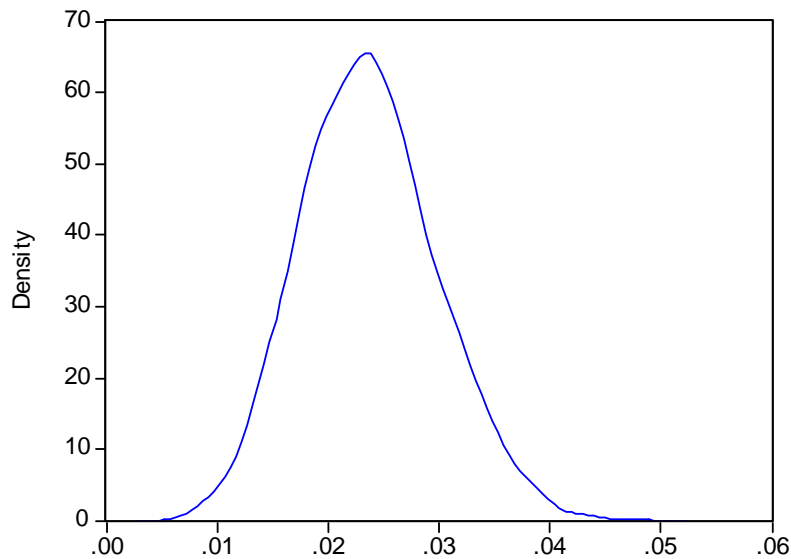


Figure 5: Posterior Predictive Density for V_{T+1} . The ‘True’ Simulated Value is $V_{T+1} = 0.015846$.

The MCMC draws of λ_{T+1} along with the those of σ_v are combined with the ‘known’ value of ρ in (7) to produce a predictive density for the relative risk aversion parameter, γ_{T+1} , corresponding to λ_{T+1} . This density is shown in Figure 7, with the true value of γ_{T+1} , 4.9908, being easily enclosed by the 95% HPD interval.

Finally, we use iterates of all relevant unknowns to produce, via (27), an estimate of the predictive of the (average) conditional risk premium for the next 22 day period beginning at day $T + 1$,¹⁵ denoted by

$$crp_{T+1} = \{a_\tau V_{T+1} + b_\tau - [a_{T+1,\tau}^* V_{T+1} + b_{T+1,\tau}^*]\} / 22.$$

Given the actual (simulated) value of realized volatility over the future 22 day period beginning at day $T + 1$, $RV_{T+1,T+1+\tau}$, and the actual (simulated) value of the option-implied measure at time $T + 1$, $MF_{T+1,T+1+\tau}$, we use the quantity, $\{RV_{T+1,T+1+\tau} - MF_{T+1,T+1+\tau}\} / 22 = -0.0187$ to evaluate predictive accuracy. As is clear, the observed value falls well within the high mass region of the predictive distribution¹⁶.

¹⁵Note that in the construction of draws from crp_{T+1} from draws of the unknowns on which crp_{T+1} depends, we use draws of λ_{T+1} . An implication of this is that the risk premium parameter is held fixed over the 22 day future period to which the conditional risk premium relates. In this sense the risk premium is conditional on *both* the prediction of V_{T+1} and the prediction of λ_{T+1} .

¹⁶Note: the predictive density for the observable quantity $RV_{T+1,T+1+\tau} - MF_{T+1,T+1+\tau}$ itself would appropriately cater for the measurement error in both $MF_{T+1,T+1+\tau}$ and $RV_{T+1,T+1+\tau}$ and would, as a

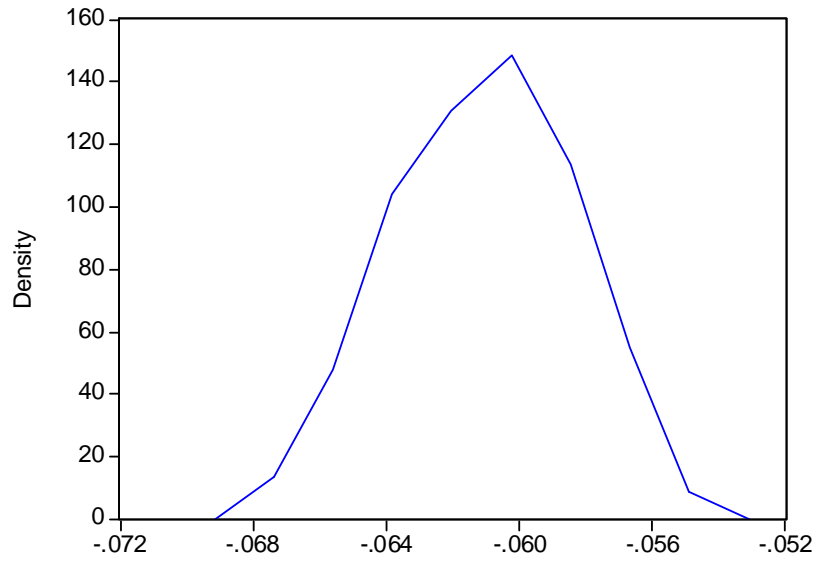


Figure 6: Posterior Predictive Density for λ_{T+1} . The ‘True’ Simulated Value is $\lambda_{T+1} = -0.05989$.

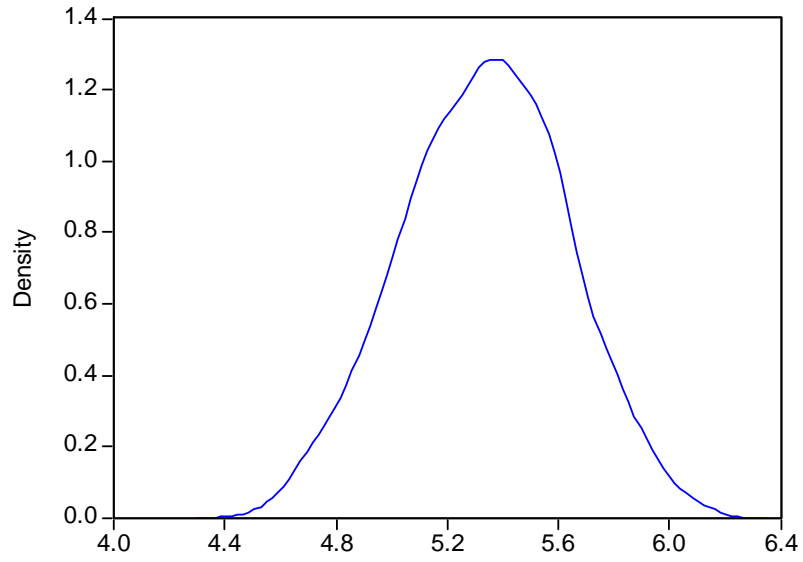


Figure 7: Posterior Predictive Density for $\gamma_{T+1} = \frac{\lambda_{T+1}}{\sigma_v \rho}$, Based on $\rho = -0.4$. The ‘True’ Value of γ_{T+1} is 4.9908.

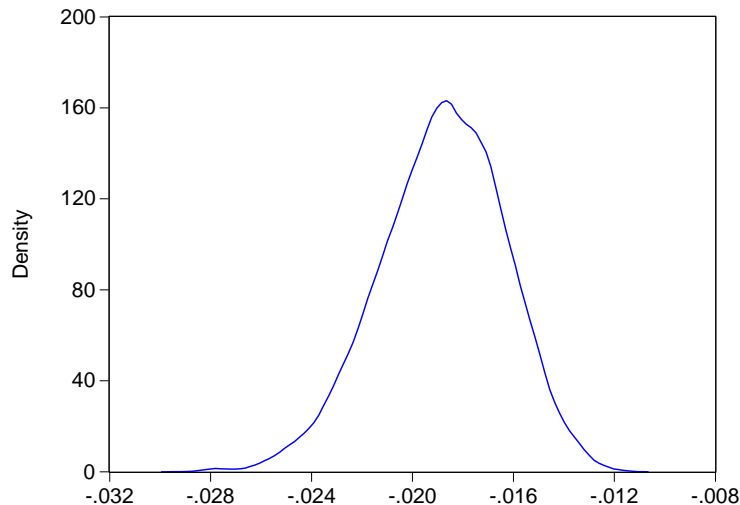


Figure 8: Posterior Predictive Density for crp_{T+1} . The Observed Value of $RV_{T+1, T+1+\tau} - MF_{T+1, T+1+\tau}$ is -0.0187 .

6 Empirical Investigation

6.1 Data Description

In this section we report results of the application of the algorithm to intraday spot and option price data for the S&P500 index from November 8, 2004 to 30 May 2006. A more extended version of this data set has been examined recently in a comprehensive forecast evaluation exercise conducted by Martin et al. (2009). All index data has been supplied by the Securities Industries Research Centre of Asia Pacific (SIRCA) on behalf of Reuters, with the raw index data having been cleaned using the methods of Brownlees and Gallo (2006). The calculations of the model free option-implied measure are based on the implied volatility surface data provided by IVOLATILITY (www.ivolatility.com). This surface comprises implied volatilities for options on the index with values of moneyness (K/P_t) ranging from 0.5 to 1.5 in steps of 0.1, and with varying times to maturity. The raw option data from which the surface is constructed is end-of-day out-of-the-money (OTM) put and call quote data.¹⁷

A realized volatility measure based on fixed 5 minute sampling is used in the analysis.

consequence, be more diffuse than the predictive shown here, which caters simply for parameter uncertainty and uncertainty in future λ_t and V_t .

¹⁷The Black-Scholes model is used to produce the implied volatilities given that options written on the S&P500 index are European. More details on the construction of the surface are available at http://www.ivolatility.com/doc/IVolatility_Data_detailed.pdf.

The measure is based on artificial returns five minutes apart with an interpolation method used to construct these returns. Note that the various forms of microstructure noise-adjusted measures that have appeared in the literature have their prime motivation in the case of data on traded assets, rather than observations on a constructed index. However, one could argue that the presence of stale prices in the index at the point of any recorded up-date induces a form of noise. With this view we use a subsampled (or averaged) version of the 5-minute based measure as an additional form of noise adjustment (i.e. in addition to sampling the observations at fixed 5 minute intervals).

Details of the calculation of $MF_{t,t+\tau}$ in (32) are as follows. Given maximum and minimum strike values K_{\max} and K_{\min} respectively,

$$\begin{aligned} MF_{t,t+\tau} &= 2 \int_{X_{\min}}^{X_{\max}} \frac{C(t+\tau, K)e^{r_t\tau} - \max\left[0, P_t^{(D)}e^{r_t\tau} - K\right]}{K^2} dK \\ &\approx \sum_{j=1}^G [g(T, K_j) + g(T, K_{j-1})] \Delta K, \end{aligned} \quad (39)$$

where $\Delta K = (K_{\max} - K_{\min})/G$, $K_j = K_{\min} + j\Delta K$ for $0 \leq j \leq G$, $g(T, K_j) = (C(t+\tau, K_j)e^{r_t\tau} - \max\left[0, P_t^{(D)}e^{r_t\tau} - K_j\right])/K_j^2$, $P_t^{(D)}$ = the (dividend-discounted) spot index at time t and r_t = the (annualized) risk free rate of return at time t . Given the finite number of points ($< G$) on the moneyness spectrum of the IVOLATILITY surface, the approach of Jiang and Tian (2005) is adopted, with steps as follows: 1) Extract the IVOLATILITY one-month (22 trading day) implied volatilities for moneyness values in the range: $0.5 < K/P_t < 1.5$ in steps of 0.1¹⁸; 2) Use linear interpolation between these values to produce a smooth function of implied volatilities and use this function to extract implied volatilities at the G grid points K_j ; 3) Use the Black-Scholes model to translate the K_j into ‘observed’ prices $C(t+\tau, K_j)$;¹⁹ 4) Use the full set of G K_j and $C(t+\tau, K_j)$ values to calculate $MF_{t,t+\tau}$ as in (39). In Martin et al. (2009) it is demonstrated that $MF_{t,t+\tau}$, as constructed from a slightly truncated moneyness range, is indistinguishable from the publicly available *VIX*

¹⁸Note that this curve itself has been produced via an initial interpolation procedure given the quoted option prices for particular strikes.

¹⁹The Black-Scholes (BS) option price model assumes that the asset price, P_t , follows a geometric Brownian motion process with constant diffusion parameter σ . Under this distributional assumption, the BS price of a European call option with strike price K and maturity $t+\tau$ is $BS = P_t^{(D)}\Phi(d_1) - Ke^{-r_t\tau}\Phi(d_2)$ where $d_1 = \left(\ln(P_t^{(D)}/K) + (r_t + 0.5\sigma^2)\tau\right)/\sigma\sqrt{\tau}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, and $\Phi(\cdot)$ = the cumulative normal distribution. An observed market option price at time t for a call option with maturity $t+\tau$ and strike K , $C(t+\tau, K)$, can be used to produce an estimate of σ implied by $C(t+\tau, K)$, by equating $C(t+\tau, K)$ to the right-hand-side of the expression for BS and solving for σ . As noted by Jiang and Tian (2005), the BS model is only being used as a mechanism to produce (artificially) a larger range of option prices than is available in practice, with this curve fitting procedure not requiring the BS model to be the ‘true’ model underlying the observed prices.

measure constructed by the Chicago Board Option Exchange (CBOE) using the model-free methodology. See the CBOE website (www.cboe.com) for more details.

In order to assess the impact of using jumps-adjusted measures in (31) both (32), the bi-power measure of Barndorff-Nielsen and Shephard (2004) is also used. Defining *realized bi-power variation* as

$$BPV_t = \frac{\pi}{2} \sum_{t_{i-1}, t_i \in [t-1, t]} |r_{t_i}| |r_{t_{i-1}}|, \quad (40)$$

where r_{t_i} denotes the i th return and $t_{i-1}, t_i \in [t-1, t]$ denote the time points at which prices are recorded on day t . Barndorff-Nielsen and Shephard show that as the number of transactions $\rightarrow \infty$, $BPV_t \xrightarrow{p} \mathcal{V}_{t-1, t}$ even in the presence of jumps in the process for P_t the integrated variance of the continuous sample path component of the price process in (1).²⁰ This is in contrast with RV_t , which is a consistent estimator of the sum of the continuous sample path and jump variation, in the presence of jumps. Hence, we re-produce the empirical results using BPV_t in (31), in replacement of RV_t , in order to see whether the use of a measure which does not contain jump information alters the inferences drawn about the continuous sample path variation that is being explicitly modelled.

As noted earlier, Jiang and Tian (2005) demonstrate that the equality in (16) holds under general jump-diffusion processes, in which case the option-based measure in (32) incorporates (in part) information about the risk-neutral expectation of jump variation over the maturity period of the option. In the spirit of Bollerslev, Gibson and Zhou (2008), we approximate this information by:

$$c[RV_t - BPV_t]$$

for two different values of c , $c = 0.2, 0.5$. The measure in (32) is then replaced by the ‘jump adjusted’ measure,

$$MF_{t, t+\tau}^a = MF_{t, t+\tau} - c[RV_t - BPV_t]$$

and the impact on the results to the change of measure documented.

In Figure 9 we reproduce, respectively plots of RV_t and $MF_{t, t+\tau}$, BPV_t and $MF_{t, t+\tau}^a$ ($c = 0.2$) and BPV_t and $MF_{t, t+\tau}^a$ ($c = 0.5$). The empirical regularity of the option-implied variance exceeding (in the main) the realized variance is in evidence in Panel A. Despite the option-implied measure being much less noisy than the returns-based measure, both measures exhibit broadly similar fluctuations, with there being only a slight tendency for the peaks in

²⁰Analogous to the adjustment made to RV_t we implement an averaged (or subsampled) version of BV_t . Following Andersen, Bollerslev and Diebold (2005), and Huang and Tauchen (2005), we also replace the sum of absolute adjacent returns in (40) with the sum of the corresponding one-period staggered returns, as a further attempt to mitigate ‘noise’ in the recorded index values. More details can be found in Martin et al. (2009).

$MF_{t,t+\tau}$ to lag those in RV_t . The ‘jump-free’ bi-power measure, BVP_t , plotted in Panels B and C has a lower average magnitude than the corresponding RV_t , as well as being smoother in appearance. Similarly, the two adjusted option-implied measures ($MF_{t,t+\tau}^a(c = 0.2)$ in Panel B and $MF_{t,t+\tau}^a(c = 0.5)$ in Panel C) are lower in magnitude than the ‘raw’ measure in Panel A. However, only $MF_{t,t+\tau}^a(c = 0.5)$ exhibits a markedly smoother appearance. $MF_{t,t+\tau}^a(c = 0.5)$ is also very close in magnitude to BVP_t at many time points.

6.2 Empirical Results

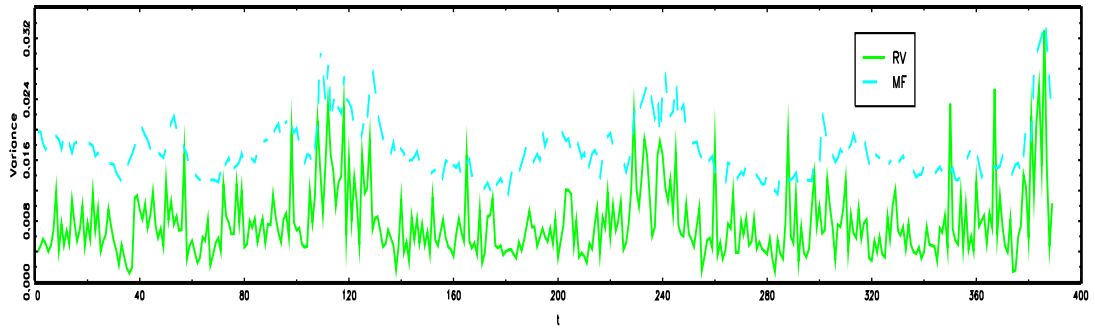
In Tables 2 and 3 we report the results based on estimation of the model in (31) to (34), with the simplifications noted at the beginning of Section 4 imposed. We record the results based on RV_t and $MF_{t,t+\tau}$ in Table 2 and the corresponding results for BVP_t and $MF_{t,t+\tau}^a$ in Table 3, for the two different values of c . All results indicate a reasonably high level of persistence in both the latent volatility process and the process for λ_t . As might be anticipated, once the option-implied and returns-based measures are both adjusted for jumps (in particular for the case where the option-implied measure is calculated using a high reduction ($c = 0.5$) for jump variation, the degree of persistence estimated in the continuous sample path volatility increases. Overall, however, the numerical results regarding the given model are reasonably robust to adjustments made to the measures for the jumps that are not being explicitly modelled.

In addition to recording point and interval estimates of the fixed parameters, we produce summaries of the one-step-ahead forecast distributions for the latent variables themselves, $V_{T+1} = V_{390}$ and $\lambda_{T+1} = \lambda_{390}$, for the risk aversion parameter, $\gamma_{T+1} = \gamma_{390}$ and for $crp_{T+1} = crp_{390}$. The forecasts of $\lambda_{T+1} = \lambda_{390}$ are (as anticipated) all negative, but with the value for λ_t resulting from estimation with $MF_{t,t+\tau}^a(c = 0.5)$ and BVP_t being 50% smaller (in magnitude) than that produced by the $MF_{t,t+\tau}$ and RV_t measures. The value for γ_{390} is forecast using values of $\rho = -0.4, -0.8$. The forecasts range from approximately 10 to 30, depending on both the measures used in the estimation procedure and the value adopted for ρ . This spread of values is somewhat consistent with the broad range of estimates - produced via very different means - that have been reported for this parameter in the literature (see Cochrane, 2005, for some recent discussion).

Finally, the predictions of latent volatility also vary depending on the measures used, with the $MF_{t,t+\tau}^a/BVP_t$ -based forecasts slightly smaller than the $MF_{t,t+\tau}/RV_t$ -based forecast. In the case of the RV_t -based results, the 95% prediction interval encompasses the relevant observed value of objective volatility (namely $RV_t = 0.0149$). For the BVP_t results however, the two intervals exclude the observed value $BVP_t = 0.014171$. Of course, the predictive for the latent variable is *not* equivalent to the predictive for the measurement. In particular,

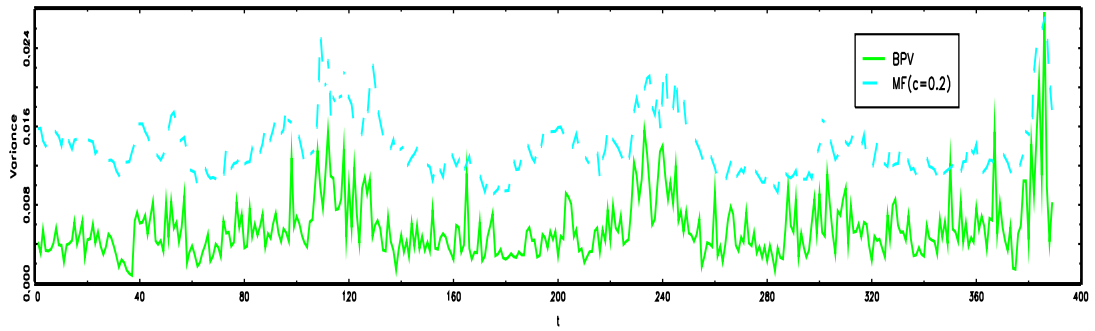
Panel A: Daily Realized Volatility and Option-Implied Volatility

S&P500 Index: Nov 2004 to May 2006



Panel B: Daily Bi-Power Variation and Option-Implied Volatility: Jump-Adjusted ($c = 0.2$)

S&P500 Index: Nov 2004 to May 2006



Panel C: Daily Bi-Power Variation and Option-Implied Volatility: Jump-Adjusted ($c = 0.5$)

S&P500 Index: Nov 2004 to May 2006

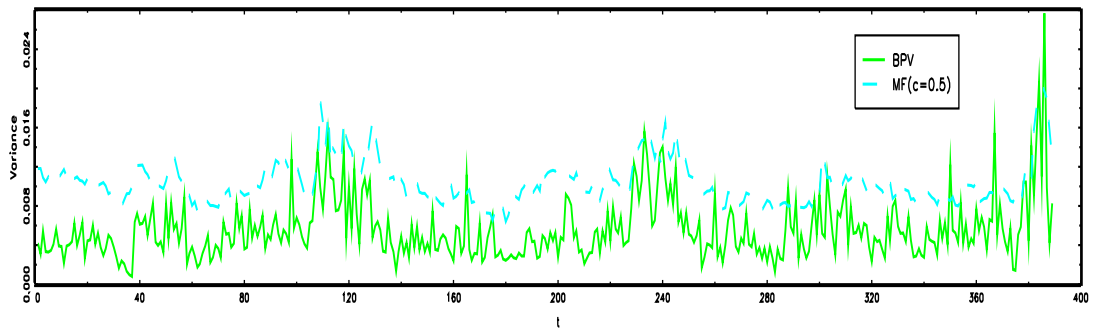


Figure 9: Daily Variance Measures for the S&P500 Index: 8 November 2004 to 30 May 2006

Table 2: Empirical Results for the S&P500 Stock Index for November 2004 to May 2006. Marginal Posterior Estimates Based on 15000 Draws Following a 35000 Burn-in Period.

Measures: RV_t and $MF_{t,t+\tau}$		
Parameter	MPM	95% <i>HPD</i>
κ	0.16607	(0.14811, 0.17982)
θ	0.00614	(0.00565, 0.00674)
σ_v	0.01768	(0.01520, 0.02029)
σ_{RV}	0.00376	(0.00346, 0.00407)
σ_{MF}	0.00133	(0.00113, 0.00154)
$\alpha_0 (1 - [\alpha_1 - \delta])$	-0.1594	(-0.16944, -0.14703)
$\alpha_1 + \delta$	0.66017	(0.23964, 0.99962)
Forecast values		
V_{390}	0.01119	(0.00684, 0.01561)
λ_{390}	-0.1594	(-0.1696, -0.1470)
$\gamma_{390} (\rho = -0.4)$	22.6619	(19.2693, 26.1727)
$\gamma_{390} (\rho = -0.8)$	11.3310	(9.6346, 13.0864)
crp_{390}	-0.0136	(-0.0165, -0.0107)

if we were to correctly incorporate the measurement error for BPV_t in the production of a predictive density for this measure itself, the observed value at $t = 390$ may well fall within the relevant 95% bounds. Our primary interest, however, is in prediction of the latent V_t and in common with all volatility forecast evaluation exercises, it is a moot point as to which observed measure should be used to assess predictive accuracy. The most thorough approach would be to assess our predictions of V_t using a range of different possible objective measures, including measures other than the one used in the estimation exercise, and to compare the results with those relevant to alternative forecasts of V_t . If our density forecast of V_t were deemed to be more accurate overall (via prediction interval coverage; log scoring; probability integral transform (PIT) assessment etc.) than its competitors, no matter what observed measure were used in the assessment, then its superiority would be confirmed. This more complete type of assessment is left for future work.

Table 3: Empirical Results for the S&P500 Stock Index for November 2004 to May 2006. Jump-Adjusted Measures Used. Marginal Posterior Estimates Based on 15000 Draws Following a 35000 Burn-in Period.

Parameter	Measures: BPV_t and $MF_{t,t+\tau}^a$			
	$c = 0.2$		$c = 0.5$	
	MPM	95% <i>HPD</i>	MPM	95% <i>HPD</i>
κ	0.20988	(0.19938, 0.22152)	0.12069	(0.10412, 0.14000)
θ	0.00404	(0.00377, 0.00433)	0.00493	(0.00439, 0.00546)
σ_v	0.01724	(0.01511, 0.01965)	0.01291	(0.01055, 0.01546)
σ_{RV}	0.00236	(0.00215, 0.00257)	0.00230	(0.00212, 0.00249)
σ_{MF}	0.00116	(0.00098, 0.00133)	0.00082	(0.00071, 0.00094)
$\alpha_0 (1 - [\alpha_1 - \delta])$	-0.20454	(-0.21270, -0.19636)	-0.10487	(-0.11358, -0.09466)
$\alpha_1 + \delta$	0.66400	(0.25042, 0.99995)	0.60550	(0.18290, 0.99984)
<hr/>				
Forecast values				
V_{390}	0.00792	(0.00436, 0.01152)	0.00848	(0.00562, 0.01129)
λ_{390}	-0.2046	(-0.2127, -0.1964)	-0.10512	(-0.11378, -0.09441)
$\gamma_{390} (\rho = -0.4)$	29.8017	(25.8968, 33.7981)	20.5425	(16.5842, 24.5498)
$\gamma_{390} (\rho = -0.8)$	14.9009	(12.9484, 16.8991)	10.27125	(8.2921, 12.2749)
crp_{390}	-0.01159	(-0.0143, -0.0090)	-0.00687	(-0.0084, -0.0054)

7 Conclusions

This paper is the first to combine non-Gaussian state space techniques with Bayesian inferential methodology for the purpose of isolating objective volatility from its time-varying risk premium, using both option and returns-based volatility measures. Whilst the work is still at a reasonably preliminary stage, the results are very encouraging, with the methodology producing (for artificially simulated data) very accurate estimates of the parameters of the dynamic volatility process and reasonably accurate estimates of those associated with the dynamic risk premium parameter. The method has been applied to 18 months of daily data on the S&P500 index, with both ‘raw’ and ‘jump-adjusted’ measures having been used to produce numerical results.

References

- [1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. 2003. Modelling and Forecasting Realized Volatility, *Econometrica*, 71: 579-625.
- [2] Barndorff-Nielsen, O.E. and Shephard, N. 2002. Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society B*, 64: 253-280.
- [3] Barndorff-Nielsen, O.E. and Shephard, N. 2004. Power and Bipower Variation with Stochastic Volatility and Jumps, *Journal of Financial Econometrics*, 2: 1-37.
- [4] Bates, D.S. 2000. Post-87 Crash Fears in the S&P 500 Futures Option Market, *Journal of Econometrics*, 94: 181-238.
- [5] Black, F. and Scholes, M. 1973. The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 81: 637-659.
- [6] Blair, B.J, Poon, S-H. and Taylor, S.J. 2001. Forecasting S&P100 Volatility: the Incremental Information Content of Implied Volatilities and High Frequency Index Returns, *Journal of Econometrics*, 105, 5-26.
- [7] Bollerslev, T., Gibson M. and Zhou, H. 2008. Dynamic Estimation of Volatility Risk Premia and Investor Risk Aversion from Option-Implied and Realized Volatilities, *Working Paper, Division of Research and Statistics, Federal Reserve Board*.
- [8] Bollerslev, T., Sizova, N. and Tauchen, G. 2009. Volatility in Equilibrium: Asymmetries and Dynamic Dependencies, *Working paper, School of Economics and Management, University of Aarhus, Denmark*.
- [9] Bollerslev, T. and Zhou, H. 2002. Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility, *Journal of Econometrics*, 109, 33-65.
- [10] Breeden, D.T. 1979. An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities, *Journal of Financial Economics*, 7, 265-296.
- [11] Britten-Jones, M. and Neuberger, A. 2000. Option Prices, Implied Price Processes and Stochastic Volatility. *The Journal of Finance*, LV: 839-866.
- [12] Carr, P. and Wu, L. 2004. Variance Risk Premia, *Working Paper*.

- [13] Carter, C.K. and Kohn, R. 1994. On Gibbs Sampling for State Space Models, *Biometrika*, 81, 541-553.
- [14] 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.
- [15] Cochrane, J.H. 2005. *Asset Pricing*, Second Edition, Princeton University Press.
- [16] Corradi, V and Swanson, N. 2006. Predictive density and conditional confidence interval accuracy tests, *Journal of Econometrics*, 135, 187–228.
- [17] Cox, J.C., J.E. Ingersoll and Ross, S.A. 1985. A Theory of the Term Structure of Interest Rates, *Econometrica*, 53, 385-408.
- [18] Creal, D.D. 2008. Analysis of Filtering and Smoothing Algorithms for Levy-driven Stochastic Volatility Models, *Computational Statistics and Data Analysis*, 52, 2863-2876.
- [19] Diebold, FX, Gunther, T and Tay, A. 1998. Evaluating Density Forecasts with Applications to Financial Risk Management, *Int Econ Rev*, 39, 863-883.
- [20] Engle, R.F. and Russell, J.R. 1998. Autoregressive Conditional Duration: A New Approach for Irregularly Spaced Transaction Data, *Econometrica*, 66, 987-1007.
- [21] Eraker, B. 2004. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices.” *The Journal of Finance*, LIX: 1367-1403.
- [22] Feigen, P.D., Gould, P., Martin, G.M. and Snyder, R.D. 2008. Feasible Parameter Regions for Alternative Discrete State Space Models, *Statistics and Probability Letters*, 78, 2963-2970.
- [23] Forbes C.S., Martin, G.M. and Wright J. 2007. Inference for a Class of Stochastic Volatility Models Using Option and Spot Prices: Application of a Bivariate Kalman Filter, *Econometric Reviews, Special Issue on Bayesian Dynamic Econometrics*, 26: 387-418.
- [24] Gneiting, T, Balabdaoui, F and Raftery, A. 2007. ‘Probabilistic forecasts, calibration and sharpness’, *JRSS (B)*, 69, 243–268.
- [25] Guo, D. 1998. The Risk Premium of Volatility Implicit in Currency Options, *Journal of Business and Economic Statistics*, 16: 498-507.

- [26] Hansen, P.R. 2005. A Test for Superior Predictive Ability, *Journal of Business and Economic Statistics*, 23: 365-380.
- [27] Hansen, P.R. and Lunde, A. 2005. A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?, *Journal of Applied Econometrics*, 20: 873-889.
- [28] Heinen, A. (2003) Modelling Time Series Count Data: an Autoregressive Conditional Poisson Model, *Draft Paper, University of California, San Diego*.
- [29] 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.
- [30] Hull, J.C. and White, A. 1987. The Pricing of Options on Assets with Stochastic Volatilities, *Journal of Finance*, 42, 281-300.
- [31] Jiang, G.J. and Tian, Y.S. 2005. The Model-Free Implied Volatility and its Information Content, *The Review of Financial Studies*, 18: 1305-1342.
- [32] Johannes, M., Polson, N.G. and Stroud, J.R. 2008. Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices, *Working Paper*
- [33] Jones, C. 2003. The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets, *Journal of Econometrics*, 116, 181-224.
- [34] Jung, R.C., Kukuk, M. and Leisenfeld, R. (2006) Time Series of Count Data: Modeling, Estimation and Diagnostics, *Computational Statistics and Data Analysis*, 51, 2350-2364.
- [35] Koopman, S. J., Jungbacker, B. and Hol, E. 2005. Forecasting Daily Variability of the S&P100 Stock Index using Historical, Realized and Implied Volatility Measurements, *Journal of Empirical Finance*, 12: 445-475.
- [36] Martens, M. and Zein, J. 2004. Predicting Financial Volatility: High-Frequency Time Series Forecasts Vis-a-Vis Implied Volatility, *Journal of Futures Markets*, 24: 1005-1028.
- [37] Martin, G.M., Reidy, R. and Wright, J. 2008. Does the option market produce superior forecasts of noise-corrected volatility measures? *Journal of Applied Econometrics*, 24, 77-104.
- [38] Neely, C.J. 2003. Forecasting Exchange Volatility: is Implied Volatility the Best we Can Do? *Working Paper, Federal Reserve Bank of St. Louis*.

- [39] Pan, J. 2002. The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-series Study, *Journal Of Financial Economics*, 63, 3-50.
- [40] Polson, N. G. and Stroud, J.R.. 2003. Bayesian Inference for Derivative Prices, *Bayesian Statistics*, 7, 641-650.
- [41] Pong, S., Shackleton, M.B., Taylor, S.J. and Xu, X. 2004. Forecasting Currency Volatility: a Comparison of Implied Volatilities and AR(FI)MA Models, *Journal of Banking and Finance*, 28: 2541-2563.
- [42] Stroud, J.R., Muller, P. and Polson, N.G. 2003. Nonlinear State-space Models with State-Dependent Variances, *Journal of the American Statistical Association*, 98, 377-386.
- [43] Wu, L. 2005. Variance Dynamics: Joint Evidence from Options and High Frequency Returns, *Draft Paper*.

Appendix A: Generation of $V|\phi, MF, RV$ Defining the two-dimensional measurement vector,

$$y_t = \begin{bmatrix} RV_t \\ MF_t \end{bmatrix},$$

the model in (31) to (34) (with $f(V_t) = \sigma_{RV}^2$) can be represented in state space form as

$$y_t = c_t + Z_t V_t + u_t, \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \sim N(0_2, H_t) \quad (41)$$

$$V_{t+1} = d_t + R_t V_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q_t) \quad (42)$$

where

$$c_t = \begin{bmatrix} 0 \\ \tau\theta_t^* - \frac{\theta_t^*}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) \end{bmatrix}, \quad Z_t = \begin{bmatrix} 1 \\ \frac{1}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) \end{bmatrix}$$

$$d_t = \kappa\theta\Delta t, \quad R_t = (1 - \kappa\Delta t),$$

and

$$H_t = \begin{bmatrix} \sigma_{RV}^2 & 0 \\ 0 & \sigma_{MF}^2 \end{bmatrix} \quad Q_t = \sigma_v^2 \Delta t V_t.$$

For simplicity, we will assume that the initial state $V_1 \sim N(\theta, \sigma_1^2)$. (For example, we could set $\sigma_1 = \sigma_v^2 \theta / (2\kappa)$, so that the mean and variance of this convenient form are those of the marginal distribution of the stochastic variance process.)

From this state space representation, and ignoring the conditioning on other fixed parameters in the model, the joint density of the unobserved stochastic variances, V , the observed realized volatilities, RV , and the model-free implied volatilities, MF , may be represented as the product of the initial state density, the $T - 1$ remaining state transition densities and the T measurement densities,

$$p(RV, MF, V) = p(V_1) p(y_1|V_1) \prod_{t=2}^T [p(V_t|V_{t-1}) p(y_t|V_t)] \quad (43)$$

where the density functions are given by

$$p(V_1) = (2\pi)^{-1/2} \sigma_1^{-1} \exp\left\{-\frac{1}{2\sigma_1^2} (V_1 - \theta)^2\right\},$$

$$p(V_t|V_{t-1}) = (2\pi)^{-1/2} Q_{t-1}^{-1/2} \exp\left\{-\frac{1}{2Q_{t-1}} (V_t - d_{t-1} - R_{t-1}V_{t-1})^2\right\}, \quad (44)$$

for $t = 2, 3, \dots, T$, and

$$p(y_t|V_t) = (2\pi)^{-1} |H_t|^{-1/2} \exp\left\{-\frac{1}{2} (y_t - c_t - Z_t V_t)' H_t^{-1} (y_t - c_t - Z_t V_t)\right\},$$

for $t = 1, 2, \dots, T$.

To deal with the complication of the state variable, V_t , appearing in the state variance, Q_t , we use an MH algorithm, with the candidate model based on an application the mixture-model approach developed by Stroud *et al.* (2003). Specifically, the approximating mixture model is obtained by replacing the state equation in (42) by a constant variance Gaussian model,

$$V_{t+1} = d_t + R_t V_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q[z_t]),$$

where

$$Q[z_t] = \sigma_v^2 \Delta t V[z_t],$$

and the set $\{V[1], V[2], \dots, V[K]\}$ represents a grid of K points in the support of the latent stochastic variances. The grid value is selected according to

$$z_t = k \text{ with probability } p^a \left(z_t = k | Y, V_t^{(ig-1)} \right),$$

where Y denotes the $(T \times 2)$ matrix of observations with t^{th} row $y_t' = (RV_t, MF_t)$. That is, the candidate draw \tilde{V} is made, conditional upon the previously sampled stochastic variance vector, $V^{(ig-1)}$, by first drawing²¹

$$\tilde{Z} \sim p^a \left(Z | Y, V^{(ig-1)} \right),$$

where $Z = \{Z_t, t = 1, 2, \dots, T\}$ and subsequently drawing \tilde{V} from the T -dimensional Gaussian distribution

$$\tilde{V} \sim p^a \left(V | Y, \tilde{Z}, V^{(ig-1)} \right).$$

Focusing initially on drawing \tilde{Z} , we note that

$$p^a(Z|Y, V) \propto p^a(Y|V, Z) p^a(Z|V).$$

Conditional upon V , the observed y_t are mutually independent, with

$$p^a(Y|V, Z) = \prod_{t=1}^T p^a(y_t | V_t, Z).$$

Note, however, that in this somewhat simplified setting, the bivariate measurement equation is linear and Gaussian. Hence, there is no need to modify the form of the conditional distribution of $p(y_t | V_t)$ in the approximation, and so $p^a(y_t | V_t, Z) \equiv p(y_t | V_t)$. This density is thus unaffected by the value of z_t , and so

$$p^a(Z|V, Y) \propto p^a(Z|V).$$

²¹The superscript ‘ a ’ distinguishes the ‘approximating’ densities from the true densities.

In the spirit of Stroud *et al.* (2003), the mixture indicator probabilities are determined from the past draw of volatilities, $V^{(ig-1)}$, according to independent multinomial distributions. The individual probabilities are selected using a Gaussian kernel, centred at each of the K grid points, and having scale parameter set equal to some fixed value, σ_{v^a} . That is, the multinomial probabilities are given by

$$p^a \left(z_t = k | V_t^{(ig-1)} \right) = C \left(V_t^{(ig-1)} \right)^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{V_t^{(ig-1)} - V[k]}{\sigma_{v^a}} \right)^2 \right\},$$

where

$$C \left(V_t^{(ig-1)} \right) = \sum_{l=1}^K \exp \left\{ -\frac{1}{2} \left(\frac{V_t^{(ig-1)} - V[l]}{\sigma_{v^a}} \right)^2 \right\}.$$

Conditional upon \tilde{Z} and Y , the approximating model for the vector of latent stochastic variances is a linear, Gaussian state space model, and hence the Forward Filter Backwards Sampling (FFBS) algorithm (e.g. Carter and Kohn, 1994), may be used to generate \tilde{V} . This conditional distribution,

$$\tilde{V} \sim p^a \left(V | Y, \tilde{Z}, V^{(ig-1)} \right) \equiv p^a \left(V | Y, \tilde{Z} \right),$$

has density function proportional to the product of the approximating initial state density, the approximating state transition densities and the measurement densities

$$p^a \left(V | Y, \tilde{Z}, V^{(ig-1)} \right) \propto p^a \left(V_1 | \tilde{z}_1 \right) p \left(y_1 | V_1 \right) \prod_{t=2}^T p^a \left(V_t | V_{t-1}, \tilde{z}_t \right) p \left(y_t | V_t \right).$$

Marginalising over the choice of \tilde{z}_t yields

$$\begin{aligned} p^a \left(V_t | V_{t-1}, V^{(ig-1)} \right) &= \sum_{k=1}^K p^a \left(V_t | V_{t-1}, z_t = k \right) p^a \left(\tilde{z}_t = k | V_t^{(ig-1)} \right) \\ &= \sum_{k=1}^K \left[(2\pi)^{-\frac{1}{2}} Q[k]^{-1/2} \exp \left\{ -\frac{1}{2Q[k]} \left(V_t - d_{t-1} - R_{t-1} V_{t-1} \right)^2 \right\} \right. \\ &\quad \left. \times C \left(V_t^{(ig-1)} \right)^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{V_t^{(ig-1)} - V[k]}{\sigma_{v^a}} \right)^2 \right\} \right], \end{aligned} \quad (45)$$

for $t = 2, 3, \dots, T$. Here $Q[k] = \sigma_v^2 \Delta t V[k]$. Hence, the approximating density for \tilde{V} given Y (and marginal of \tilde{Z}) is given by

$$p^a \left(V | Y, V^{(ig-1)} \right) \propto p^a \left(V_1 | V_1^{(ig-1)} \right) p \left(y_1 | V_1 \right) \prod_{t=2}^T p^a \left(V_t | V_{t-1}, V_t^{(ig-1)} \right) p \left(y_t | V_t \right),$$

or

$$\begin{aligned}
p^a(V|Y, V^{(ig-1)}) &\propto p^a(V_1|V_1^{(ig-1)}) p(y_1|V_1) \\
&\times \prod_{t=2}^T p(y_t|V_t) \sum_{k=1}^K \left[(2\pi)^{-\frac{1}{2}} Q[k]^{-1} \exp \left\{ -\frac{1}{2Q[k]} (V_t - d_{t-1} - R_{t-1}V_{t-1})^2 \right\} \right. \\
&\left. \times C(V_t^{(ig-1)})^{-1} \exp \left\{ -\frac{1}{2} \left(\frac{V_t^{(ig-1)} - V[k]}{\sigma_{v^a}} \right)^2 \right\} \right]. \tag{47}
\end{aligned}$$

The density of the initial state, $p^a(V_1|V_1^{(ig-1)})$, is taken as a degenerate distribution at $V_1 = \theta$.

Finally, we write the Metropolis-Hastings acceptance/rejection step as: accept $V^{(ig)} = \tilde{V}$ with probability

$$a(V^{(ig-1)}, \tilde{V}) = \min \left\{ 1, \frac{p(\tilde{V}|Y)}{p(V^{(ig-1)}|Y)} / \frac{p^a(\tilde{V}|Y, V^{(ig-1)})}{p^a(V^{(ig-1)}|Y, \tilde{V})} \right\}. \tag{48}$$

The target density, $p(V|Y)$, is available up to a constant of proportionality in (43) whereas the candidate density from the approximation, $p^a(\tilde{V}|Y, V^{(ig-1)})$, is (47). Substituting into the ratio in (48), and reducing common factors in the numerator and denominator, the Metropolis-Hasting ratio is

$$\begin{aligned}
\frac{p(\tilde{V}|Y)}{p(V^{(ig-1)}|Y)} / \frac{p^a(\tilde{V}|Y, V^{(ig-1)})}{p^a(V^{(ig-1)}|Y, \tilde{V})} &= \frac{p(\tilde{V}|Y)}{p(V^{(ig-1)}|Y)} \times \frac{p^a(V^{(ig-1)}|Y, \tilde{V})}{p^a(\tilde{V}|Y, V^{(ig-1)})} \\
&= \prod_{t=2}^T \frac{p(\tilde{V}_t|\tilde{V}_{t-1}) p^a(V_t^{(ig-1)}|V_{t-1}^{(ig-1)}, \tilde{V}_t)}{p(V_t^{(ig-1)}|V_{t-1}^{(ig-1)}) p^a(\tilde{V}_t|\tilde{V}_{t-1}, V_t^{(ig-1)})}.
\end{aligned}$$

The transition densities from the target model are given in (44), and the transition densities

from the approximating model, $p^a(V_t|V_{t-1})$, are given in (46), so that

$$\begin{aligned}
& \frac{p(\tilde{V}_t|\tilde{V}_{t-1})p^a(V_t^{(ig-1)}|V_{t-1}^{(ig-1)}, \tilde{V}_t)}{p(V_t^{(ig-1)}|V_{t-1}^{(ig-1)})p^a(\tilde{V}_t|\tilde{V}_{t-1}, V_t^{(ig-1)})} \\
&= \frac{\tilde{V}_{t-1}^{-1/2} \exp\left\{-\frac{1}{2\sigma_v^2\Delta t\tilde{V}_t}(\tilde{V}_t - d_{t-1} - R_{t-1}\tilde{V}_{t-1})^2\right\}}{\left(V_t^{(ig-1)}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma_v^2\Delta tV_t^{(ig-1)}}(V_t^{(ig-1)} - d_{t-1} - R_{t-1}V_{t-1}^{(ig-1)})^2\right\}} \\
&\times \frac{C(\tilde{V}_t)^{-1} \sum_{k=1}^K Q[k]^{-1/2} \exp\left\{-\frac{1}{2Q[k]}(V_t^{(ig-1)} - d_{t-1} - R_{t-1}V_{t-1}^{(ig-1)})^2\right\} \exp\left\{-\frac{1}{2}\left(\frac{\tilde{V}_t - V[k]}{\sigma_v^a}\right)^2\right\}}{C(V_t^{(ig-1)})^{-1} \sum_{k=1}^K Q[k]^{-1/2} \exp\left\{-\frac{1}{2Q[k]}(\tilde{V}_t - d_{t-1} - R_{t-1}\tilde{V}_{t-1})^2\right\} \exp\left\{-\frac{1}{2}\left(\frac{V_t^{(ig-1)} - V[k]}{\sigma_v^a}\right)^2\right\}}.
\end{aligned}$$

Appendix B: Generation of $\phi|V, MF, RV$ After specifying initial values ($i = 0$) for the parameters in the vector ϕ , the steps for the Gibbs chain for iteration $i = i + 1$ are as follows:

1. Sample the parameter function $\lambda_c = \alpha_0/(1 - \alpha_1 - \delta)$ via a random walk MH algorithm with tuning (variance) parameter $\varsigma_\lambda = 0.0005$.
2. Sample the parameters α_1 and δ jointly via a bivariate random walk MH algorithm, tuning variance parameters $\varsigma_\alpha = 0.05$ and $\varsigma_\delta = 0.005$ and covariance parameter $\varsigma_{\alpha,\delta} = -0.75$.
3. Sample κ via an MH algorithm with candidate distribution as detailed in (56) below.
4. Sample θ via an MH algorithm with candidate distribution derived analogously to that of κ .
5. Sample σ_V directly from its full conditional, as detailed in (57) below.
6. Sample σ_{MF} directly from its full conditional, as detailed in (58) below.
7. Sample σ_{RV} directly from its full conditional, as detailed in (59) below.

The full conditional of κ is given by

$$\begin{aligned}
p(\kappa | \phi_{-\kappa}, MF, RV) &\propto \left[\prod_{t=2}^T p(MF_{t,t+\tau} | MF_{t-1,t-1+\tau}, RV_{t-1}, V_t, \kappa, \theta, \sigma_{MF}, \alpha_0, \alpha_1, \delta)\right] \\
&\times p(V_t | V_{t-1}, \kappa, \theta, \sigma_v) p(\kappa),
\end{aligned}$$

where $\phi_{-\kappa}$ denotes all parameters in ϕ other than κ , and the prior, $p(k)$, encompasses all of the restrictions that the model imposes on κ , namely:

$$\begin{aligned}\kappa &> 0 \\ \kappa + \lambda_t &> 0 \text{ for all } t \\ \sigma_V^2 &\leq 2\kappa\theta.\end{aligned}$$

As this conditional is non-standard, we define a candidate distribution as follows. Denoting by \overline{MF}_t , the conditional mean of $MF_{t,t+\tau}$ in (32), we take a first order Taylor Series expansion of

$$\overline{MF}_t = \tau\theta_t^* - \frac{\theta_t^*}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) + \frac{1}{\kappa_t^*} (1 - e^{-\tau\kappa_t^*}) V_t \quad (49)$$

about the previous draw of κ (denoted here by κ^p) as

$$\widehat{MF}_t = \overline{MF}_t(\kappa^p) + \overline{MF}_t'(\kappa^p)(\kappa - \kappa^p), \quad (50)$$

where

$$\overline{MF}_t' = \frac{1}{\kappa_t^{*2}} (1 - e^{-\kappa_t^* \tau}) \left(\frac{\theta_t^*}{2} - \theta - V_t \right) + \frac{1}{\kappa_t^*} \tau e^{-\kappa_t^* \tau} (V_t - \theta_t^*) + \frac{\tau}{\kappa_t^*} (\theta - \theta_t^*). \quad (51)$$

The parameter κ enters the expression for κ_t^* in (49) and (51) via the relationship:

$$\begin{aligned}\kappa_t^* &= \kappa + \lambda_t \\ &= \kappa + \alpha_0 + \alpha_1 \lambda_{t-1} + \delta l_{t-1}\end{aligned} \quad (52)$$

where l_{t-1} is a non-linear function of κ defined by (36) (for period $t-1$). The parameter κ in turn affects θ_t^* through $\theta_t^* = \frac{\theta\kappa}{\kappa_t^*}$, with κ_t^* as given in (52). All instances of κ in (49) and (51) are evaluated at κ^p as per (50). Given the approximation in (50), the candidate density for κ is given by

$$\begin{aligned}p_c(\kappa \mid \phi_{-\kappa}, MF, RV) &\propto \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_{MF}^2}} \exp \left\{ -\frac{1}{2\sigma_{MF}^2} \left(MF_t - \widehat{MF}_t \right)^2 \right\} \\ &\times \frac{1}{\sqrt{2\pi V_{t-1} \Delta t \sigma_V^2}} \exp \left\{ -\frac{1}{2V_{t-1} \Delta t \sigma_V^2} \left(V_t - \overline{V}_t \right)^2 \right\}, \quad (53)\end{aligned}$$

where and \overline{V}_t is the conditional mean of V_t in (33),

$$\overline{V}_t = \kappa\theta\Delta t + (1 - \kappa\Delta t) V_{t-1}.$$

Defining Σ as a $(T - 1 \times T - 1)$ diagonal matrix with diagonal elements $V_{t-1}\Delta t$, and

$$\begin{aligned}
MF^* &= (MF_2, MF_3 \dots, MF_T)' \\
\widehat{MF} &= (\widehat{MF}_2, \widehat{MF}_3 \dots, \widehat{MF}_T)' \\
\overline{MF} &= (\overline{MF}_2, \overline{MF}_3 \dots, \overline{MF}_T)' \\
\overline{MF}' &= (\overline{MF}'_2, \overline{MF}'_3 \dots, \overline{MF}'_T)' \\
V &= (V_2, V_3 \dots, V_T)' \\
V_{-1} &= (V_1, V_2 \dots, V_{T-1})' \\
\overline{V} &= (\overline{V}_2, \overline{V}_3 \dots, \overline{V}_T)',
\end{aligned}$$

(53) can, in turn, be rewritten as follows (where the dependence of (49) and (51) on k^p is made explicit),

$$\begin{aligned}
p_c(\kappa \mid \phi_{-\kappa}, MF, RV) &\propto \exp \left\{ -\frac{1}{2\sigma_{MF}^2} \left(MF - \widehat{MF} \right)' \left(MF - \widehat{MF} \right) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \left(V - \overline{V} \right)' \Sigma^{-1} \left(V - \overline{V} \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma_{MF}^2} \left(-2MF' \widehat{MF} + \widehat{MF}' \widehat{MF} \right) \right\} \\
&\times \exp \left\{ -\frac{1}{2} \left(-2V' \Sigma^{-1} \overline{V} + \overline{V}' \Sigma^{-1} \overline{V} \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2\sigma_{MF}^2} \left[-2MF' \overline{MF}'(\kappa^p) \kappa + \overline{MF}(\kappa^p)' \overline{MF}'(\kappa^p) \right. \right. \\
&\quad \left. \left. \times \kappa + \overline{MF}'(\kappa^p)' \overline{MF}'(\kappa^p) (\kappa^2 - 2\kappa \kappa^p) \right] \right\} \tag{54} \\
&\times \exp \left\{ -\frac{1}{2} \left(-2V' \Sigma^{-1} (\theta - V_{-1}) \kappa \Delta t + (\theta - V_{-1})' \Sigma^{-1} (\theta - V_{-1}) \kappa^2 \Delta t^2 \right) \right\} \\
&\propto \exp \left\{ -\frac{1}{2} (c_{\kappa^2} \kappa^2 + c_{\kappa} \kappa) \right\} \\
&\propto \exp \left\{ -\frac{1}{2 c_{\kappa^2}^{-1}} \left(\kappa - \frac{-c_{\kappa}}{2c_{\kappa^2}} \right)^2 \right\}. \tag{56}
\end{aligned}$$

This defines a Gaussian density with mean $\frac{-c_{\kappa}}{2c_{\kappa^2}}$ and variance $c_{\kappa^2}^{-1}$ where

$$c_{\kappa^2} = \frac{1}{\sigma_{MF}^2} \left(\overline{MF}'(\kappa^p)' \overline{MF}'(\kappa^p) \right) + (\theta - V_{t-1})' \Sigma^{-1} (\theta - V_{t-1}) \Delta t^2$$

and

$$c_{\kappa} = 2 \left(\frac{1}{\sigma_{MF}^2} \left(-MF' \overline{MF}'(\kappa^p) + \overline{MF}'_t(\kappa^p)' \overline{MF}'(\kappa^p) - \overline{MF}'(\kappa^p)' \overline{MF}'(\kappa^p) \kappa^p \right) - V' \Sigma^{-1} (\theta - V_{t-1}) \Delta t \right)$$

The full conditional of σ_V has density

$$p(\sigma_V | \phi_{-V}, MF, RV) \propto \frac{1}{(\sigma_V^2)^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} (V - \bar{V})' \Sigma^{-1} (V - \bar{V}) \right\} p(\sigma_V) \quad (57)$$

where ϕ_{-V} denotes all parameters in ϕ other than σ_V , and the prior, $p(\sigma_V)$, encompasses the restriction that the model imposes on σ_V , namely,

$$\sigma_V^2 \leq 2\kappa\theta.$$

The density in (57) corresponds to a an inverse gamma kernal with shape parameter $\frac{T}{2}$ and scale parameter $\frac{1}{2} (V - \bar{V})' \Sigma^{-1} (V - \bar{V})$.

The full conditional of σ_{MF} has density

$$p(\sigma_{MF} | \phi_{-MF}, MF, RV) \propto \frac{1}{(\sigma_{MF}^2)^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2\sigma_{MF}^2} (MF - \overline{MF})' (MF - \overline{MF}) \right\} p(\sigma_{MF}) \quad (58)$$

where ϕ_{-MF} denotes all parameters in ϕ other than σ_{MF} , and the prior, $p(\sigma_{MF})$ is uniform (truncated from below at zero). The density in (58) corresponds to a an inverse gamma kernal with shape parameter $\frac{T}{2}$ and scale parameter $\frac{1}{2} (MF - \overline{MF})' (MF - \overline{MF})$. Similarly, the full conditional of σ_{RV} has density

$$p(\sigma_{RV} | \phi_{-RV}, MF, RV) \propto \frac{1}{(\sigma_{RV}^2)^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2\sigma_{RV}^2} (RV_t - V_t)' (RV_t - V_t) \right\} p(\sigma_{RV}) \quad (59)$$

where ϕ_{-RV} denotes all parameters in ϕ other than σ_{RV} , and the prior, $p(\sigma_{RV})$ is uniform (truncated from below at zero). The density in (59) corresponds to a an inverse gamma kernal with shape parameter $\frac{T}{2}$ and scale parameter $-\frac{1}{2} (RV_t - V_t)' (RV_t - V_t)$.