Deviance Information Criterion as a Model Comparison Criterion for Stochastic Volatility Models *

Andreas Berg,† Renate Meyer,† and Jun Yu‡

February 14, 2002

Abstract

Bayesian methods have proven very efficient in estimating parameters of stochastic volatility (SV) models for analysing financial time series. Recent work extends the basic stochastic volatility model to include heavy-tailed error distributions, covariates, leverage effects, and jump components. Hierarchical Bayesian methods (usually implemented via state-of-the-art Markov chain Monte Carlo methods for posterior computation) allow fitting of such complex models. However, a formal model comparison via Bayes factors is difficult because the marginalization constants are not readily available. Bayesian model comparison using the Schwarz criterion as a Bayes factor approximation requires the specification of the number of free parameters in the model. This number of free parameters, or degrees of freedom, is not well defined in stochastic volatility models. The main objective of this paper is to demonstrate that model selection within the class of SV models is better performed using the deviance information criterion (DIC). DIC is a recently developed information criterion designed for complex hierarchical models with possibly improper prior distributions. It combines a measure of fit with a measure of model complexity. We illustrate the performance of DIC in discriminating between various different SV models using simulated data and daily returns data on the S&P 100 index.

*Andreas Berg gratefully acknowledges the support of his research by the DAAD, personal number D/99/19367 within a "DAAD Doktorandenstipendium im Rahmen des gemeinsamen Hochschulsonderprogramms III von Bund und Laendern". The work of Renate Meyer was supported by the Royal Society of New Zealand Marsden Fund and the University of Auckland Research Committee. Jun Yu would like to acknowledge financial support from the University of Auckland Research Committee and the Marsden Fund under No. 01-UOA-015. We acknowledge comments from Adrian Pagan, Werner Ploberger and seminar participants at the 2001 Australasian Econometric Society Meeting in Auckland.

†Department of Statistics, University of Auckland

‡Department of Economics, University of Auckland

Keywords: *Stochastic Volatility, Model Selection, Deviance, Model Complexity, Gibbs Sampler, BUGS, MCMC, Leverage Effect, Heavy-tailed Distributions, Jumps, Implied Volatility.*

1 Introduction

The progress in Bayesian posterior computation due to Markov chain Monte Carlo (MCMC) methods has made it possible to fit increasingly complex statistical models and entailed the wish to determine the best-fitting model in a potentially huge class of candidates. Thus, it has become more and more important to develop efficient model selection criteria. A recent proposal by Spiegelhalter *et al.* (2001) is the *Deviance Information Criterion* (DIC), a Bayesian version or generalization of the well-known Akaike Information Criterion (AIC), related also to the Bayesian Information Criterion (BIC). DIC is easy to calculate and applicable to a wide range of statistical models. It is based on the posterior distribution of the log-likelihood or the deviance, following the original suggestion of Dempster (1974) for model choice in the Bayesian framework. It overcomes the problem of identifying the number of parameters in the models which is required for the calculation of BIC and AIC. This is accomplished by defining the *effective number of parameters* in a complex hierarchical Bayesian model. This model comparison criterion has already been successfully applied to complex models in the field of medical statistics (Zhu *et al.* (2000)). In this paper, we demonstrate its usefulness in the model selection process for financial time series. The aim of this paper is therefore to introduce DIC to the financial modelling community and show how to use it for the family of stochastic volatility (SV) models.

Indeed, many model checking criteria (Gelman *et al.* (1996), Gilks *et al.* (1996), Key *et al.* (1999), Carlin and Louis (1996)) have been proposed and discussed before the development of DIC. While Bayes factors (e.g. Kass and Raftery (1995)) have been viewed for many years as the only correct way to carry out Bayesian model comparison, they have come under increasing criticism of late (Kass and Raftery (1995), Lavine and Schervish (1999)). The most serious drawback is that they are not well-defined when using improper priors which is typically the case in practice when employing noninformative priors. This led to modifications, such as the *partial Bayes factor* (O'Hagan, 1991), the *intrinsic Bayes factor* (Berger and Pericchi, 1996), and the *fractional Bayes factor* (O'Hagan, 1994). These modifications suffer from more or less arbitrary choices of training samples, weights for averaging training samples, and fractions, respectively. Furthermore, the calculation of Bayes factors can be computationally difficult due to an implicit high-dimensional integration problem. As such, the computation of Bayes factors is not a particularly user-friendly tool for practising statisticians. In their review of MCMC methods for computing Bayes factors, Han and Carlin (2001, page 29) conclude that *'all of the methods ... discussed require substantial time and effort (both human and computer) for a rather modest payoff, namely a collection of posterior model probability estimates ... As a result, one might conclude that none of the methods herein is appropriate for everyday, "rough and ready" model comparison, and instead search for more computationally realistic alternatives'*.

Shortcuts to the calculation of Bayes factors that avoid multidimensional integration through the large sample approximations of $-2\log(\text{Bayes-factor})$ include the familiar BIC, also referred to as *Schwarz Criterion* (Schwarz, 1978), and the related penalized likelihood ratio model choice criterion, AIC. Both criteria demand the specification of the number of parameters. In hierarchical Bayesian models, however, the number of unknowns often exceeds the number of observations and thus, neither BIC nor AIC is applicable.

Chib *et al.* (2001) and Kim *et al.* (1998) use the marginal likelihood approach of Chib (1995) to calculate Bayes factors from the likelihood ordinates at arbitrary points (recommended at the posterior means) using the technique of particle filtering.

As this requires normal error distributions, they convert the multivariate stochastic volatility model into a collection of Gaussian state space models, showing that the log-volatilities of the SV model follow approximately a seven component mixture of normal distributions. This approach is based on approximations and is computationally intensive. The above-mentioned technique of Chib (1995) is also recommended by Han and Carlin (2000) for a wide range of hierarchical problems, but limitations of this method are outlined for models of increasing complexity.

The outline of the paper is as follows: Section 2 starts with a short description of SV models, followed in Section 3 by the introduction of DIC. In Section 4, we present results from a simulation study. Section 5 shows how DIC is applied to compare the fit of various SV models to a dataset previously used in the literature. Section 6 concludes.

2 The Stochastic Volatility Model

In the literature on the financial volatility, the SV model (Tauchen and Pitts (1983) and Taylor (1982)) has received much attention in recent years, mainly because it uses the time-varying variability of the volatility to describe the dynamics of financial data such as interest rates, exchange rates, and stock market prices. It has become a powerful alternative to the ARCH and GARCH models (Engle (1982), Bollerslev (1986)).

A basic, simple stochastic volatility model consists of an observation equation

$$y_t = \exp(\frac{h_t}{2}) \cdot u_t, \qquad t = 1, \dots, n,$$

that describes the distribution of the data given unknown states, the volatilities, and a state equation

$$h_t = \mu + \phi \cdot (h_{t-1} - \mu) + \sigma \cdot \eta_t, \qquad t = 1, \dots, n,$$

which models the day-to-day variation of the volatilities as a Markov process. Here, y_t is the response variable with $u_t \overset{iid}{\sim} N(0, 1)$, and h_t is the log-volatility process with η_t

as a white noise. We collect the three model parameters in a vector $z = (\phi, \mu, \sigma)$. In section 4 we will introduce extensions of the basic model to more complex SV models. An example of such an extension is the inclusion of a level effect in the observation equation, namely

$$y_t = x_t^\gamma \cdot exp(\frac{h_t}{2}) \cdot u_t, \qquad t = 1, \ldots, n,$$

where x_t denotes a time-varying covariate. The parameter γ plays an important role in analyzing interest rate data (for details refer for example to Chan *et al.* (1992) and Brenner *et al.* (1996)). In other applications, for example stock market data, it is common to set this parameter equal to 0 (see Section 5 below).

The SV model is a typical example of a hierarchical model, in which the number of unknowns, i.e. the parameters (z) and the unknown states (h_t) exceeds the number of observations. The number of free parameters in the model could be claimed as the number of model parameters (3) or the number of states plus the number of model parameters (n+3) or something in between. In any case this number is not well defined. Classical parameter estimates for this model is extremely difficult, because of the non-analytic form of the likelihood function. Harvey *et al.* (1994) employed a quasi-maximum likelihood technique, whereas Sandmann and Koopman (1998) used maximum likelihood Monte Carlo. Several method of moment approaches like efficient method of moments (Gallant and Tauchen (1996) and Andersen *et al.* (1999)), simulated method of moments (Duffie and Singleton (1993)) or generalized method of moments (Melino and Turnbull (1990)) have also been used to estimate the model parameters.

Whilst some of the above mentioned techniques use adhoc criteria[1], a Bayesian approach is based on a sound statistical paradigm. Bayesian posterior computations are performed using MCMC techniques. Several different algorithms have been proposed

[1] See Andersen *et al.* (1999) for a review and comparison of various estimation techniques for the SV model.

by Jacquier *et al.* (1994), Kim *et al.* (1998) and further developed in Chib *et al.* (1999, 2001). Although more efficient updating techniques for SV models exist,[2] for ease of implementation we employ the all purpose Bayesian software package BUGS based on the single-update Gibbs sampler. We will show that DIC provides an efficient and straightforward approach to identifying the most appropriate model.

3 The Deviance Information Criterion (DIC)

Assume, in general, that the conditional distribution of the data, $y = (y_1, ..., y_n)$, depends on a p-dimensional parameter vector θ. From a frequentist point of view, model assessment is based on the *deviance*, the difference in the log-likelihoods between the fitted and the saturated model. The saturated model refers to the model with as many parameters as observations, that yields a perfect fit to the data. By analogy, for Bayesian model selection, Dempster (1974) suggested to examine the posterior distribution of the classical deviance defined by

$$D(\theta) = -2 \log f(y \mid \theta) + 2 \log f(y).$$

Here, $\log f(y)$ serves as a standardizing term. Dempster (1974) suggested plots and potential summaries such as the posterior mean of $D(\theta)$ and Spiegelhalter *et al.* (2001) followed these suggestions in the development of DIC as a model choice criterion. Based on the posterior distribution of $D(\theta)$, DIC consists of two components: a term that measures goodness-of-fit and a penalty term for increasing model complexity. The measure of fit consists of the posterior expectation of the deviance

$$E_{\theta|y}[D] = \bar{D}.$$

The 'better' the model fits the data, the larger are the values for the likelihood. $\bar{D}$ which is defined via -2 times log-likelihood therefore attains smaller values for 'better'

[2]For latest developments with regard to different specific SV models we refer to Kim *et al.* (1998) and Chib *et al.* (2001).

models.

The second component measures the complexity of the model by the *effective number of parameters*, p_D, defined as the difference between posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta}) = E_{\theta|y}[-2\log f(y \mid \theta) + 2\log f(y \mid \bar{\theta}(y))].$$

From the latter expression it can be seen that p_D can be regarded as the expected excess of the true over the estimated residual information in data y conditional on θ. Alternatively we can interpret p_D as the expected reduction in uncertainty due to estimation.

Spiegelhalter *et al.* (2001) also examine the properties of p_D as a penalty term for approximately normal likelihoods and negligible prior information, a situation which adequately mimics a frequentist point of view. In this situation, the posterior mean equals the maximum likelihood (ML) estimate $\hat{\theta}$ and the posterior distribution of θ is well approximated by a multivariate normal distribution with mean equal to the ML estimate

$$\theta \mid y \approx N(\hat{\theta}, -\frac{\partial^2\theta}{\partial\theta^2}|_{\theta=\hat{\theta}}).$$

Representing the deviance through a second order Taylor expansion around the ML estimate, the first order term vanishes and the deviance is approximated by the sum of the deviance at the ML estimate and a quadratic term that follows a χ^2 distribution with p degrees of freedom. By taking expectations the effective number of parameters equals p, the true number of parameters.

DIC is then defined as the sum of both components

$$\text{DIC} = \bar{D} + p_D = 2\bar{D} - D(\bar{\theta}) = D(\bar{\theta}) + 2p_D.$$

In a "classical" situation, as described above, DIC can be seen as a generalization of the well-known AIC, defined by AIC $= D(\hat{\theta}) + 2p$ where $\hat{\theta}$ is the ML estimate of the parameter vector and p the number of parameters. Thus for non-hierarchical models, $p \approx p_D$, $\hat{\theta} \approx \bar{\theta}$, and DIC $\approx$ AIC.

By applying a logarithmic loss function, Spiegelhalter *et al.* (2001) show that DIC approximately describes the expected posterior loss when adopting a particular model. For additional asymptotic properties of p_D and $\bar{D}$ the interested reader is referred to Spiegelhalter et al. (2001). In striking contrast to Bayes factors and BIC, computing DIC via MCMC is almost trivial. p_D and thus DIC, is easily calculated from an MCMC output by taking the sample mean of the simulated values of $D(\theta)$, minus the plug-in estimate of the deviance using the sample means of the simulated values of θ (see also section 4.3). So far, no efficient method has been developed for calculating reasonably accurate MC standard errors of DIC. Zhu *et al.* (2000) explore this problem, but their approach using the multivariate delta method yields poor results. Their final recommendation is the "brute force" approach, which is simply replicating the calculation of DIC some N times and estimating VAR(DIC) by its sample variance

$$\widehat{\text{VAR}}(\text{DIC}) = \frac{1}{N-1}\sum_{k=1}^{N}(\text{DIC}_k - \overline{\text{DIC}})^2.$$

Although this is a painfully time-consuming approach, it at least gives an indication of the inherent variability of DIC.

4 A Simulation Study

We simulated a dataset comprising 2000 observations from a stochastic volatility+ Jumps model with parameters as described below. The selected SV+Jumps model (SVJ Model 1) is very similar to the one proposed in the simulation analysis by Chib *et al.* (2001). We then applied DIC as a model choice criterion. For a posterior analysis of our models, we used the BUGS package which is available free of charge via internet from http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.

BUGS is an easy to learn and easy to use Bayesian software package that implements the Gibbs sampler for generating samples from a Markov chain whose equilibrium distribution is the posterior distribution. As demonstrated by Meyer and Yu (2000), it can be applied to fit stochastic volatility models.

4.1 The Models

We fitted seven different stochastic volatility models to the simulated data. The first five models are identical to those in Meyer and Yu (2000) and follow their notation. Additionally we consider two slightly different SV+Jumps models, one of them (SVJ Model 1) is the correct model with which the simulation was carried out.

SV Model 1:

$$\begin{aligned} y_t \mid h_t &= \exp(\tfrac{1}{2} \cdot h_t) \cdot u_t, & u_t &\overset{iid}{\sim} N(0,1),\ t = 1, ..., n, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2 &= \mu + \phi \cdot (h_{t-1} - \mu) + v_t, & v_t &\overset{iid}{\sim} N(0, \tau^2), \end{aligned}$$

with $h_0 \sim N(\mu, \tau^2)$. SV Model 1 is identical to the basic stochastic volatility model in Section 2.

SV Model 2: An additional non-zero mean α is added in the observation equation:

$$\begin{aligned} y_t \mid h_t, \alpha &= \alpha + \exp(\tfrac{1}{2} \cdot h_t) \cdot u_t, & u_t &\overset{iid}{\sim} N(0,1),\ t = 1, ..., n, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2 &= \mu + \phi \cdot (h_{t-1} - \mu) + v_t, & v_t &\overset{iid}{\sim} N(0, \tau^2). \end{aligned}$$

SV Model 3: An AR(2)-process for the state equation:

$$\begin{aligned} y_t \mid h_t &= \exp(\tfrac{1}{2} \cdot h_t) \cdot u_t, & u_t &\overset{iid}{\sim} N(0,1), \quad t = 1, ..., n, \\ h_t \mid h_{t-1}, \mu, \phi, \psi, \tau^2 &= \mu + \phi \cdot (h_{t-1} - \mu) + \psi \cdot (h_{t-2} - \mu) + v_t, & v_t &\overset{iid}{\sim} N(0, \tau^2). \end{aligned}$$

SV Model 4: A central Student-t distribution with ν degrees of freedom, for the observation error term:

$$\begin{aligned} y_t \mid h_t &= \exp(\tfrac{1}{2} \cdot h_t) \cdot u_t, & u_t &\overset{iid}{\sim} t_\nu,\ t = 1, ..., n, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2 &= \mu + \phi \cdot (h_{t-1} - \mu) + v_t, & v_t &\overset{iid}{\sim} N(0, \tau^2). \end{aligned}$$

SV MODEL 5: The basic SV MODEL 1 including a leverage or asymmetric effect implemented by allowing for correlation between u_t and v_{t+1}, i.e.

$$\begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \overset{iid}{\sim} N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right\},\ h_0 \sim N(\mu, \tau^2).$$

This effect is often observed in financial time series, e.g. in time series of exchange rates and, even stronger, in stock market data. It reveals the market behavior, first discovered by Black (1976), which "occurs when an unexpected drop in price (bad news) increases predictable volatility more than an unexpected increase in price (good news) of similar magnitude" (Engle and Ng (1993, page 1752)).

SVJ MODEL 1: The SV+Jumps (SVJ) model 1 includes a jump component in the observation equation and a Gaussian error distribution:

$$\begin{aligned} y_t \mid h_t, s_t, q_t, b, \alpha &= \alpha + b \cdot y_{t-1} + s_t \cdot q_t + \exp(\frac{1}{2} \cdot h_t) \cdot u_t, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2, &= \mu + \phi \cdot (h_{t-1} - \mu) + v_t, \end{aligned}$$

where q_t follows a Bernoulli distribution which takes the value of one with probability κ and zero with probability $1 - \kappa$, and s_t (the jump size when a jump takes place) follows the distribution

$$s_t \sim N(0, \delta^2).$$

SVJ MODEL 2: The SV+Jumps (SVJ) model 2 includes a jump component in the observation equation but without taking the lagged observations into consideration:

$$\begin{aligned} y_t \mid h_t, s_t, q_t, \alpha &= \alpha + s_t \cdot q_t + \exp(\frac{1}{2} \cdot h_t) \cdot u_t, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2 &= \mu + \phi \cdot (h_{t-1} - \mu) + v_t. \end{aligned}$$

4.2 Prior Distributions

We basically use the same prior distributions suggested by Kim *et al.* (1998). The prior for parameter μ is a normal distribution with mean 0 and variance 10. A conjugate inverse gamma distribution with parameters 2.5 and 0.025 was chosen for τ^2. Defining $\phi = 2\phi^* - 1$, we specify a beta-distribution with parameters 20 and 1.5 for ϕ^*. For α (SV Model 2) a flat normal distribution with mean parameter $\mu_\alpha = 0$ and variance $\sigma_\alpha^2 = 10$ was specified. We use the same prior for ψ as for the AR(1) parameter ϕ. As the value for ϕ diminishes dramatically for SV Model 3, we depart from the beta-distribution for ϕ^* and choose a more general uniform (0,1)- distribution for ϕ^* and ψ^* respectively.

In SV Model 4, the additional parameter ν, representing the unknown degree of freedom of the assumed t-distribution, has a prior χ^2-distribution with 8 degrees of freedom. The correlation parameter ρ (SV Model 5) is assumed uniformly distributed with support between -1 and 1. As the parameter b in SVJ Model 1 is assumed to be small a priori, we use a noninformative normal distribution with hyperparameters $\mu_b = 0$ and $\sigma_b^2 = 0.04$. The parameter q_t represents the frequency of a jump occurance with a Bernoulli distribution with parameter κ. Following a prior assumption of discovering on average one jump in approximately every 10th observation, we specify a beta-distribution with parameters 2.9 and 26.1 for κ.

We assume a normal prior distribution for the jump size s_t according to our mean guess around 0. A conjugate inverse gamma distribution with parameters 25 and 2.5 is used for modelling the scale parameter δ^2.

4.3 Implementation in WinBUGS

WinBUGS is the BUGS-version operating under WINDOWS. A DIC module which automatically calculates values for DIC and related parameters is implemented in the latest WinBUGS version. Even without the DIC module, DIC is easily obtained from any MCMC output.

The first part of DIC, $\bar{D}$, is easily calculated using the MCMC output $\theta^{(i)}, i =$

$1, \ldots, N$. We simply calculate $D(\theta^{(i)})$ for $i = 1, \ldots, N$ and estimate $\bar{D}$ by the sample mean $\frac{1}{N}\sum_{i=1}^{N} D(\theta^{(i)})$. In practice, using BUGS, this is accomplished by adding the variable $D(\theta)$. For the second part, the effective number of parameters p_D, we only need to evaluate $D(\theta)$ at the sample posterior mean $\bar{\theta} = \frac{1}{N}\sum_{i=1}^{N} \theta^{(i)}$. WinBUGS offers several useful convergence checking criteria available in an attached CODA (Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Best *et al.* (1995)) module running for example under SPLUS. It is necessary to check whether convergence has been achieved because it is crucial that the sample is taken from the stationary distribution. The CODA package consists of a selection of model checking criteria, one of which is the Heidelberger and Welch test (Heidelberger and Welch (1983)).

4.4 Results

In Tables 1-7 we report means and variances of both prior and posterior distributions for each of the seven models respectively. The results for SV MODELS 1-3 AND 5 are based on 12,500 iterations. After a burn-in period of 50,000 iterations and a follow-up period of 250,000, we stored every 20th iteration. Due to higher posterior correlations amongst the parameters and thus slower convergence of the Gibbs sampler in the remaining models, we chose a burn-in period of 100,000 iterations, a follow-up period of 900,000, and stored every 40th iteration. We ran each chain 6 times to obtain a brute-force estimate of the variability of DIC. All calculations were performed on a Pentium-III PC, 550 MHz, running the WinBUGS 131 version updated with the DIC tool.

Since the SVJ MODEL 1 is the correct model, the correct parameter values used for our simulation are also listed. By re-estimating the parameters of the correct model we obtain reasonably accurate results. For all the models we have not encountered any convergence problems and the MC standard errors are very small.

Table 8 shows the smallest and largest values for DIC, the number of effective parameters p_D and the goodness-of-fit $\bar{D}$, respectively, obtained for 6 runs for each of the 7 models. The most salient result from Table 8 is that the correct model provides the smallest values for DIC, as well as for the posterior mean of the deviance and the effective number of parameters. We get only slightly larger values for the SV+Jumps model without lagged observations. This is because differences between this model and the correct model are very small.

All the other models perform clearly worse. Compared with DIC values of around 4065 and 4120 respectively for the two models with jumps, the pure SV models provide DIC values of around 5000. The pure SV model with the t-distribution has DIC values of even bigger than 5200. Although the effective number of parameters is relatively small for this model (around 145 compared to 112 for the true model), it does not fit the data well, indicated by the highest value for the posterior mean of the deviance. This adds up to the highest DIC values of all the models examined.

The SV models with normal error assumptions perform quite similar with mediocre high values for $\bar{D}$ between 4426 (AR(2)-process model) and 4550 (leverage effect model). All of them have a large effective number of parameters (close to 500). Nevertheless in terms of DIC, the AR(2)-process model is preferred. Despite the largest number of effective parameters, it seems to fit the data better then the other SV models resulting in DIC values of around 4946.

As the intercept model differs from the basic model only by an additional parameter, the number of effective parameters rises by one, which leads to DIC values just one point higher. Moreover, for our simulated dataset the leverage model presents a poor alternative to the correct model, providing the second largest DIC values.

5 An Empirical Study

5.1 The Data

The dataset consists of 1512 mean-corrected daily returns of the Standard & Poors 100 index, covering the period of time between January 1993 and December 1998. The S&P100 index returns have been used often in the literature. For instance, Blair *et al.* (2001) estimate the GJR-GARCH model proposed by Glosten, et al. (1993) based on the S&P100 index returns for four different sample periods from March 1984 to December 1998, one of which is identical to that in this paper.

We also used data from the Chicago Board Options Exchange Market Volatility Index (VIX) for the same period of time as a covariate, measuring the so-called implied volatility. For a detailed explanation of the Chicago Board Options Exchange Market Volatility Index, the reader is referred to Hol and Koopman (2000) and Fleming *et al.* (1995).

5.2 The Models

In this section, we fit the models introduced in section 4 to the above data set. In addition we consider a model that includes implied volatility:

SV MODEL 6: The SVX model (Hol and Koopman, 2000) includes implied volatility as expressed by an additional covariate x_t:

$$\begin{aligned} y_t \mid h_t &= \exp\left(\frac{1}{2} \cdot h_t\right) \cdot u_t, \quad u_t \overset{iid}{\sim} N(0,1),\ t = 1, ..., n, \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2, \lambda &= \mu + \phi \cdot (h_{t-1} - \mu) + \lambda \cdot (x_t - \bar{x}) + v_t, \\ & \quad v_t \overset{iid}{\sim} N(0, \tau^2). \end{aligned}$$

The implied volatility is used in this model as an alternative source for predicting volatility and is based on calculations of option price models. The specification of the variance equation is motivated from the empirical result that implied volatilities contain useful information in forecasting future volatilities (see for example Blair *et al.* (2001)). Note that we demean the observations in vector x_t for convergence purposes.

For an economic interpretation we allow for a positive scaling factor by setting $\beta = \exp(\frac{\mu}{2})$ which is often interpreted as the instantaneous volatility. In our case, y_t are the observations from the mean-corrected return series with volatility process h_t. The parameters μ and ϕ are the intercept and the persistence in the volatility process. The h_t's are centered around μ to avoid high posterior correlations between the parameters and thus speed up convergence of the Gibbs sampler.

A priori, λ is assumed to be uniformly distributed in the interval [0,2]. Because of the same phenomenon already discovered for SV Model 3 we choose a uniform (0,1)-distribution for ϕ^*.

We also slightly modify the structure and some prior distributions of the SV+Jumps models compared with section 4. In the context of stock market data, comparisons of SV+Jumps models and pure SV models have been illustrated for example in Eraker *et al.* (2000) and Andersen *et al.* (2000) also with additional jump components in the state equation.

Chib *et al.* (2001) work with the S&P500 dataset, represented in decimals. Using decimals allows them to apply a lognormal distribution for their jump size s_t and (because of the (expected) small values for s_t) the fact that $\log(1 + s_t) \approx s_t$ for their prior specifications. In our dataset, the values are about a factor of ten higher than in Chib *et al.* (2001). Thus, we expect a larger jump size and their quite elegant log-transformation is not applicable any more. Therefore, we use a simple Gaussian prior for the jump size s_t, with mean 0 and variance δ^2, with δ^2 following an inverse Gamma-distribution with parameters 6 and 20. Using the findings in Chib *et al.* (2001) with respect to the frequency of the jump size q_t we modify our prior beta-distribution for κ according to their posterior values. Eventually the demeaned dataset allows us to neglect the intercept parameter α.

5.3 Results

In Tables 9-16 we report means and variances of both prior and posterior distributions for each of the eight models applied to the S&P100 index respectively. The estimated means and standard deviations for the parameters appear quite reasonable and comparable with previous estimates in the literature. For instance, in the basic SV model, the volatility process is estimated to be highly persistent. The estimate of the second autoregressive coefficient in SV MODEL 3 is a large positive number and consistent with results obtained by Meyer and Yu (2000) in an exchange rate series. More interestingly, it is bigger than the estimate of the first autoregressive coefficient. In SV MODEL 4 the posterior mean of ν is 8.85 and similar to the values of 7.7 and 8.9 for the S&P500 index in Sandmann and Koopman (1998) and Chib *et al.* (2001) respectively. In SV MODEL 5 the posterior mean of ρ is -0.418 with its upper interval less than zero. It suggests that the leverage effect is significant for the S&P100 index. The posterior mean of λ in SV MODEL 6 indicates that the implied volatility contains important information about the volatility process. Interestingly, allowing for the implied volatility as a covariate induces a negative posterior mean of the autoregressive coefficient in the model. This finding is similar to what was obtained in Hol and Koopman (2000) based on a S&P100 index for a different period. The parameter estimates for the SV+Jumps models provide similar results for those parameters already covered by the pure SV models.

The jump parameters, however cannot be identified. This is indicated by posterior distributions for the jump size parameter δ and the jump frequency parameter κ that stay unchanged compared to the prior distributions.

We believe that due to the small amount of jumps in the considered dataset[3], the

[3]Multiplying our number of observations, 1512, by the jump occurrence estimate in Chib *et al.*

information provided by those observations is not sufficient for identification purposes. Table 17 shows the smallest and largest values for DIC, the number of effective parameters p_D and the goodness-of-fit $\bar{D}$, respectively, obtained for 6 runs for each of the models.

The most adequate models to describe the dataset according to DIC are the implied volatility SV Model 6, and the model including the leverage effect, SV Model 5. The jump models did not perform as well as these models, but are still ranked above SV Models 1-4. Although their fit to the data seems less satisfactory than the AR(2)-process model (values for $\bar{D}$ of around 3278 for the jump models versus 3262 for the AR(2)-process), their DIC values are smaller. These smaller DIC values result from much smaller penalty terms for model complexity p_D of only around 51 compared to around 116 for the AR(2)-process model.

influence on the model performance, the values for DIC are only marginally larger then those for SV Model 1. Quite surprising is the performance of the model with Student-t distributed errors. Although adding an additional parameter, the degrees of freedom, the effective number of parameters reduces from about 80 to 50 but its DIC values are the largest among all models that we fitted. This could be explained by a good fit of outliers, but a bad fit to the bulk of the data.

The leverage and implied volatility model perform best. Even though their effective number of parameters is higher than for the other models, the posterior means of the deviance are much lower. The effective number of parameters rises to values around 115-125 in the implied volatility model. Even almost three times higher are those in the leverage effect model. Therefore, under all the examined models, we obtain overwhelming results in terms of the goodness-of-fit $\overline{D}$ for the leverage effect model with values of around 2990 it is in contrast to 3190 for the implied volatility model.

(2001) of 0.004, which we used as our prior guess, leads to a result of only 6 jumps in the entire dataset.

6 Conclusion

In this paper we have explored the practical performance of DIC as model selection criterion for comparing various stochastic volatility models. DIC is a Bayesian version of the classical deviance for model assessment. It is particularly suited to compare Bayesian models whose posterior distributions have been obtained using MCMC simulation. Similar to AIC and BIC, DIC comprises two parts, a goodness-of-fit measure, the posterior distribution of the deviance, and a penalty term, the effective number of parameters, measuring complexity. Using this concept of *effective number of parameters*, DIC can be used in complex hierarchical models where the number of unknowns often exceeds the number of observations and the number of free parameters is not well defined. This is in contrast to AIC and BIC, where the number of free parameters needs to be specified. DIC has been implemented as a tool in the BUGS software package.

We carry out a simulation study using a SV+Jumps model as the true model. Our estimation results with respect to our simulated data set are quite accurate for all the models considered. DIC clearly identifies the correct model out of 7 different alternatives. By comparing 8 different stochastic volatility models for the S&P100 index, comprising 1512 observations from 1993 to 1998, the implied volatility model turns out to be the most adequate as indicated by the smallest DIC values. The SV model including a leverage effect also performs very well, with its remarkably outstanding performance in terms of goodness-of-fit. The Monte Carlo error of DIC was fairly low for all the models, thus indicating a stable performance for model comparison purposes.

7 References

Andersen, T.G., Benzoni, L. and Lund, J. (2000): Estimating jump-diffusions for equity returns, Forthcoming, *Journal of Finance.*

Andersen, T.G., Chung, H.-J. and Sørensen B.E. (1999): Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics* **91**, 61–87.

Berger, J.O. and Pericchi, L.R. (1996): The intrinsic Bayes factor for model selection and prediction, *The Journal of the American Statistical Association*, **91**, 109–122.

Best, N., Cowles, M.K. and Vines, K. (1995): CODA Convergence diagnosis and output analysis software for Gibbs sampling output version 0.30, MRC Biostatistics Unit, Cambridge, UK.

Blair, B.J., Poon, S.-H. and Taylor, S.J. (2000): Modelling S&P 100 volatility: The information content of stock returns, Forthcoming, *Journal of Banking and Finance.*

Blair, B.J., Poon, S.-H. and Taylor, S.J. (2001): Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency returns, Forthcoming, *Journal of Econometrics.*

Bollerslev, T. (1986): Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Brenner, R.J., Harjes, R.H. and Kroner, K.F. (1996): Another look at models of the short-term interest rate, *Journal of Financial and Quantitative Analysis*, **31**, 85–107.

Carlin, B.P. and Louis, T.A. (1996): Bayes and empirical Bayes methods for data analysis, *Monographs on Statistics and Applied Probability*, **69**, Chapman and Hall, London.

Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders A.B. (1992): An empirical comparison of alternative models of the short-term interest rate, *Journal of Finance*, **47**, 1209–1227.

Chib, S. (1995): Marginal likelihood from the Gibbs output, *The Journal of the American Statistical Association*, **90**, 1313-1321.

Chib, S., Nardari, F. and Shephard, N. (1999): Analysis of high dimensional multivariate Stochastic Volatility models, unpublished paper: John M. Olin School of Business, Washington University, St. Louis.

Chib, S., Nardari, F. and Shephard, N. (2001): Markov Chain Monte Carlo methods for generalized stochastic volatility models, Forthcoming, *Journal of Econometrics.*

Dempster, A.P. (1974): The direct use of likelihood for significance testing, *Proceedings of Conference on Foundational Questions in Statistical Inference*, Department of Theoretical Statistics: University of Aarhus, 335–352.

Duffie, D. and Singleton, K.J. (1993): Simulated moments estimation of Markov models of asset prices, *Econometrica*, **61**, 929–952.

Engle, R.F. (1982): Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Engle, R.F. and Ng, V.K. (1993): Measuring and testing the impact of news on volatility, *Journal of Finance*, Vol. XLVIII, No. 5, 1749–1778.

Eraker, B., Johannes, M.S. and Polson, N.G. (2000): The impact of jumps in volatility and returns, working paper, Graduate School of Business, University of Chicago.

Fleming, J., Ostdiek, B. and Whaley, R.E. (1995): Predicting stock market volatility: a new measure, *Journal of Futures Markets*, **15**, 265–302.

Gallant, A.R. and Tauchen, G. (1996): Which Moments to Match, *Econometric Theory* **12**, 657–681.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1996): Bayesian Data Analysis, Chapman and Hall, London.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996): Markov Chain Monte Carlo in Practice, Chapman and Hall, London.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993): On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, **48**, 1779–1801.

Han, C. and Carlin, B.P. (2001): MCMC Methods for Computing Bayes Factors: A Comparative Review, Working paper, Division of Biostatistics, School of Public Health, University of Minnesota.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994): Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247–264.

Heidelberger, P. and Welch, P. (1983): Simulation run length control in the presence of an initial transient, *Operations Research*, **31**, 1109–1144.

Hol, E. and Koopman, S.J. (2000): Forecasting the variability of stock index returns with Stochastic Volatility Models and implied volatility, Tinbergen Institute Discussion paper, No. TI 2000-104/4.

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994): Bayesian analysis of stochastic volatility models,Forecasting the variability of stock index returns with Stochastic Volatility Models, *Journal of Business and Economic statistics*, **12**, 371–389.

Kass, R.E. and Raftery, A.E. (1995): Bayes factors, *The Journal of the American Sta-*

tistical Association, **90**, 773–795.

Key, J.T., Pericchi, L.R. and Smith, A.F.M. (1999): Bayesian model choice: what and why? In Bernardo, J.M and Berger, J.O. and Dawid, A.P. and Smith, A.F.M., editors, *Bayesian Statistics 6*, 343–370.

Kim, S. and Shephard, N. and Chib, S. (1998): Stochastic volatility: likelihood inference comparison with ARCH models, *Review of Economic Studies*, **65**, 361–393.

Lavine, M. and Schervish, M.J. (1999): Bayes factors: what they are and what they are not, *The American Statistician*, **53**, 119–122.

Melino, A. and Turnbull, S.M. (1990): Pricing foreign currency options with stochastic volatility, *Journal of Econometrics*, **45**, 239–265.

Meyer, R. and Yu, J. (2000): BUGS for a Bayesian Analysis of Stochastic Volatility Models, *The Econometrics Journal*, **3**, 198–215.

O'Hagan, A. (1991): Contribution to the discussion of "Posterior Bayes Factors", *Journal of the Royal Statistical Society, Series B*, **53**, 136.

O'Hagan, A. (1994): *Kendall's Advanced Theory of Statistics, Volume 2b: Bayesian Inference*, Edward Arnold, London.

O'Hagan, A. (1995): Fractional Bayes factors for model comparison (with discussion), *Journal of the Royal Statistical Society, Series B*, **57**, 99–138.

Sandmann, G. and Koopman, S.J. (1998): Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics*, **87**, 271–301.

Schwarz, G. (1978): Estimating the dimension of a model, *Annals of Statistics*, **6**, 461-464.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2001): Bayesian measures of model complexity and fit, http://www.biostat.umn.edu/ ~ brad.

Spiegelhalter, D.J., Thomas, A., Best, N.G. and Gilks, W.R. (1996): *BUGS 0.5, Bayesian inference using Gibbs sampling. Manual (version ii).* MRC Biostatistics Unit, Cambridge, UK.

Tauchen, G. and Pitts, M. (1983): The price variability-volume relationship on speculative markets, *Econometrica*, **51**, 485–505.

Taylor, S.J. (1982): Financial returns modelled by the product of two stochastic processes — a study of the daily sugar prices 1961-75. In Anderson, O.D., Editor, *Time Series Analysis: Theory and Practice, 1*, pages 203–226. North-Holland, Amsterdam.

Zhu, L. and Carlin, B. P. (2000): Comparing hierarchical Models for spatio-temporally misaligned data using the deviance information criterion, *Statistics in Medicine*, **19**, 2265–2278.

Appendix: Tables

The tables contain the parameter estimates for mean and standard deviation for every model based on the standard priors described in Section 4. The values for DIC and p_D are summarized only for the set of modified models passing the Heidelberger and Welch convergence test for all parameters.

Table 1: Parameter estimates and MC error for the true model:

$$y_t \mid h_t, s_t, q_t, b, \alpha = \alpha + b \cdot y_{t-1} + s_t \cdot q_t + exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t,\; u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\; v_t \overset{i.i.d.}{\sim} N(0, \tau^2),\; q_t \sim Bernoulli(\kappa),$$

$$s_t \sim N(0, \delta^2)$$

Correct model=SVJ Model 1						
	True	Prior		Posterior		
Parameter	Mean	Mean	SD	Mean	SD	MC standard error
μ	-1.000	0.00	3.16	-0.842	0.163	0.001629
ϕ	0.960	0.86	0.11	0.957	0.010	0.000191
τ	0.345	0.39	0.12	0.290	0.029	0.000761
δ	3.000	3.76	1.59	3.010	0.233	0.002824
κ	0.080	0.10	0.06	0.077	0.011	0.000173
b	0.100	0.00	0.20	0.088	0.015	0.000540
α	0.001	0.00	0.45	-0.0001	0.013	0.000010

Table 2: Parameter estimates and MC error for SV Model 1:

$$y_t \mid h_t = exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 1					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.585	0.082	0.000861
ϕ	0.86	0.11	0.699	0.038	0.000783
τ	0.39	0.12	0.938	0.059	0.001314

Table 3: Parameter estimates and MC error for SV Model 2:

$$y_t \mid h_t, \alpha = \alpha + exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 2					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.586	0.082	0.000989
ϕ	0.86	0.11	0.699	0.039	0.000890
τ	0.39	0.12	0.938	0.061	0.001608
α	0.00	0.45	0.0005	0.015	0.000150

Table 4: Parameter estimates and MC error for SV MODEL 3:

$$y_t \mid h_t, \alpha = \alpha + exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \psi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + \psi \cdot (h_{t-2} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 3					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.623	0.091	0.000826
ϕ	0.00	0.58	0.283	0.062	0.001120
τ	0.39	0.12	1.060	0.055	0.000975
ψ	0.00	0.58	0.421	0.072	0.001340
α	0.00	0.45	-0.0006	0.015	0.000141

Table 5: Parameter estimates and MC error for SV MODEL 4:

$$y_t \mid h_t = exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} t_\nu, \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2, \lambda = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\ v_t \overset{i.i.d.}{\sim} N(0, \tau^2),$$

SV Model 4					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.992	0.145	0.001696
ϕ	0.86	0.11	0.948	0.014	0.000354
τ	0.39	0.12	0.293	0.037	0.001233
ν	8.00	4.00	3.080	0.251	0.005622

Table 6: Parameter estimates and MC error for SV Model 5:

$$y_t \mid h_t, \rho = exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad t = 1, ..., n.\ h_{t+1} \mid h_t, \mu, \phi, \tau^2, \rho = \mu + \phi \cdot (h_t - \mu) + \tau \cdot v_t,$$

$$\text{with} \begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \overset{i.i.d.}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}, \quad h_0 \sim N(\mu, \tau^2).$$

SV Model 5					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.548	0.079	0.000709
ϕ	0.86	0.11	0.717	0.037	0.000536
τ	0.39	0.12	0.866	0.051	0.000805
ρ	0.00	0.58	-0.076	0.047	0.000409

Table 7: Parameter estimates and MC error for SVJ Model 2:

$$y_t \mid h_t, s_t, q_t, \alpha = \alpha + s_t \cdot q_t + exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\ v_t \overset{i.i.d.}{\sim} N(0, \tau^2),\ q_t \sim Bernoulli(\kappa),$$

$$s_t \sim N(0, \delta^2)$$

SVJ Model 2					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
μ	0.00	3.16	-0.820	0.157	0.001422
ϕ	0.86	0.11	0.955	0.011	0.000214
τ	0.39	0.12	0.296	0.030	0.000799
δ	3.76	1.59	3.020	0.238	0.002928
κ	0.10	0.06	0.076	0.011	0.000176
α	0.00	0.45	0.002	0.014	0.000116

Table 8: Deviance summaries for simulated data:

Model	$\overline{D}_{min}$	$\overline{D}_{max}$	p_{Dmin}	p_{Dmax}	DIC_{min}	DIC_{max}
true model	3950.7	3954.8	111.9	113.2	4063.1	4067.8
basic SV model	4502.7	4505.4	481.0	482.2	4984.9	4987.0
intercept	4503.8	4505.1	482.1	483.5	4987.0	4988.1
AR(2)-process	4425.9	4427.0	518.8	519.3	4945.3	4946.2
t-distribution	5063.7	5065.6	144.2	145.7	5209.5	5209.8
leverage effect model	4549.2	4550.2	444.4	445.0	4994.2	4994.7
jump without lagged obs.	4004.9	4007.2	112.8	113.4	4117.7	4120.3

Table 9: Parameter estimates and MC error for SV MODEL 1:

$$y_t \mid h_t = exp\left(\tfrac{1}{2} \cdot h_t\right) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 1					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.695	0.082	0.001962
μ	0.00	3.16	-0.745	0.244	0.005914
ϕ	0.86	0.11	0.980	0.009	0.000511
τ	0.12	0.05	0.167	0.031	0.002275

Table 10: Parameter estimates and MC error for SV Model 2:

$$y_t \mid h_t, \alpha = \alpha + exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 2					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.695	0.080	0.001604
μ	0.00	3.16	-0.742	0.237	0.004826
ϕ	0.86	0.11	0.980	0.008	0.000384
τ	0.12	0.05	0.170	0.028	0.001726
α	0.00	3.16	0.010	0.017	0.000217

Table 11: Parameter estimates and MC error for SV Model 3:

$$y_t \mid h_t = exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \psi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + \psi \cdot (h_{t-2} - \mu) + v_t, \quad v_t \overset{i.i.d.}{\sim} N(0, \tau^2).$$

SV Model 3					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.679	0.086	0.001199
μ	0.00	3.16	-0.790	0.261	0.003790
ϕ	0.00	0.58	0.188	0.165	0.013230
τ	0.12	0.05	0.295	0.055	0.003499
ψ	0.00	0.58	0.779	0.163	0.012970

Table 12: Parameter estimates and MC error for SV Model 4:

$$y_t \mid h_t = exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} t_\nu, \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2, \lambda = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\ v_t \overset{i.i.d.}{\sim} N(0, \tau^2),$$

SV Model 4					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.585	0.093	0.003090
μ	0.00	3.16	-1.100	0.326	0.011060
ϕ	0.86	0.11	0.991	0.005	0.000186
τ	0.12	0.05	0.108	0.021	0.001131
ν	8.00	16.00	8.85	2.31	0.122800

Table 13: Parameter estimates and MC error for SV Model 5:

$$y_t \mid h_t, \rho = exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad t = 1, ..., n.\ h_{t+1} \mid h_t, \mu, \phi, \tau^2, \rho = \mu + \phi \cdot (h_t - \mu) + \tau \cdot v_t,$$

$$\text{with} \begin{pmatrix} u_t \\ v_{t+1} \end{pmatrix} \overset{i.i.d.}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right\}, \quad h_0 \sim N(\mu, \tau^2).$$

SV Model 5					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.711	0.063	0.000687
μ	0.00	3.16	-0.690	0.177	0.001937
ϕ	0.86	0.11	0.974	0.010	0.000337
τ	0.12	0.05	0.194	0.032	0.001376
ρ	0.00	0.58	-0.415	0.088	0.001756

Table 14: Parameter estimates and MC error for SV Model 6:

$$y_t \mid h_t = exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t, \quad u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2, \lambda = \mu + \phi \cdot (h_{t-1} - \mu) + \lambda \cdot (x_t - \bar{x}) + v_t,\ v_t \overset{i.i.d.}{\sim} N(0, \tau^2),$$

SV Model 6					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.695	0.016	0.001325
μ	0.00	3.16	-0.730	0.047	0.003720
ϕ	0.00	0.58	-0.276	0.117	0.004328
τ	0.12	0.05	0.435	0.080	0.008935
λ	1.00	0.58	0.153	0.016	0.000582

Table 15: Parameter estimates and MC error for SVJ Model 1:

$$y_t \mid h_t, s_t, q_t, b = b \cdot y_{t-1} + s_t \cdot q_t + exp\ (\tfrac{1}{2} \cdot h_t) \cdot u_t,\ u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\ v_t \overset{i.i.d.}{\sim} N(0, \tau^2),\ q_t \sim Bernoulli(\kappa),$$

$$s_t \sim N(0, \delta^2)$$

SVJ Model 1					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.676	0.089	0.00173
μ	0.00	3.16	-0.802	0.273	0.00546
ϕ	0.86	0.11	0.984	0.008	0.00027
τ	0.12	0.05	0.150	0.030	0.00134
δ	1.95	0.44	1.950	0.425	0.00998
κ	0.004	0.002	0.005	0.002	0.00004
b	0.00	0.45	0.006	0.027	0.00026

Table 16: Parameter estimates and MC error for SVJ Model 2:

$$y_t \mid h_t, s_t, q_t = s_t \cdot q_t + exp\,(\tfrac{1}{2} \cdot h_t) \cdot u_t,\; u_t \overset{i.i.d.}{\sim} N(0,1), \quad t = 1, ..., n.$$

$$h_t \mid h_{t-1}, \mu, \phi, \tau^2 = \mu + \phi \cdot (h_{t-1} - \mu) + v_t,\; v_t \overset{i.i.d.}{\sim} N(0, \tau^2),\; q_t \sim Bernoulli(\kappa),$$

$$s_t \sim N(0, \delta^2)$$

SVJ Model 2					
	Prior		Posterior		
Parameter	Mean	SD	Mean	SD	MC Standard error
β	3.49	11.67	0.677	0.088	0.001548
μ	0.00	3.16	-0.796	0.268	0.004892
ϕ	0.86	0.11	0.983	0.008	0.000347
τ	0.12	0.05	0.153	0.031	0.001747
δ	1.95	0.44	1.940	0.415	0.010230
κ	0.004	0.002	0.005	0.002	0.000036

Table 17: Deviance summaries for the S&P100 data:

	Model	$\overline{D}_{min}$	$\overline{D}_{max}$	p_{Dmin}	p_{Dmax}	DIC_{min}	DIC_{max}
1	basic SV model	3307.6	3313.0	83.2	84.9	3392.5	3396.2
2	intercept	3307.8	3311.1	84.9	87.0	3394.8	3396.5
3	AR(2)-process	3260.3	3264.3	115.5	118.6	3378.9	3379.8
4	t-distribution	3350.6	3353.0	49.1	50.4	3400.7	3402.7
5	leverage effect model	2989.2	2990.7	322.7	324.9	3313.4	3314.2
6	implied volatility	3183.2	3196.5	116.2	124.2	3307.4	3312.7
7	jump with lagged obs.	3277.4	3278.9	51.0	53.7	3329.3	3331.1
8	jump without lagged obs.	3276.1	3278.3	51.5	53.0	3327.7	3330.0