2004

Deviance Information Criterion for Comparing Stochastic Volatility Models

Andreas Berg

Renate Meyer

Jun YU
Singapore Management University, yujun@smu.edu.sg

Deviance Information Criterion for Comparing Stochastic Volatility Models[1]

Andreas Berg
Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand, andreas@stat.auckland.ac.nz
Renate Meyer
Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand, meyer@stat.auckland.ac.nz
Jun Yu
Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand, j.yu@auckland.ac.nz

Abstract

Bayesian methods have been efficient in estimating parameters of stochastic volatility models for analyzing financial time series. Recent advances made it possible to fit stochastic volatility models of increasing complexity, including covariates, leverage effects, jump components and heavy-tailed distributions. However, a formal model comparison via Bayes factors remains difficult. The main objective of this paper is to demonstrate that model selection is more easily performed using the deviance information criterion (DIC). It combines a Bayesian measure-of-fit with a measure of model complexity. We illustrate the performance of DIC in discriminating between various different stochastic volatility models using simulated data and daily returns data on the S&P100 index.

Keywords: *Model Selection, Bayesian Deviance, Model Complexity, MCMC, Leverage Effect, Jumps.*

[1]The authors would like to thank an associate editor and two anonymous referees for constructive comments which have substantially improved the paper. They also acknowledge comments from Nicky Best, Genshiro Kitagawa, Angelika van der Linde, Adrian Pagan, Peter Phillips, Werner Ploberger and seminar participants at the 2001 Australasian Econometric Society Meeting, the New Zealand Econometric Study Group Meeting in Auckland, and National University of Singapore (both Department of Economics and Center for Financial Engineering. Andreas Berg gratefully acknowledges the support of his research by the DAAD, personal number D/99/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 Royal Society of New Zealand Marsden Fund under No. 01-UOA-015.

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, Best, Carlin and van der Linde (2002) is the *Deviance Information Criterion* (DIC), a Bayesian version or generalization of the well-known Akaike Information Criterion (AIC) (Akaike 1973), related also to the Bayesian (or Schwarz) Information Criterion (BIC) (Schwarz 1978). Similar to AIC and BIC, it trades off a measure of model adequacy against a measure of complexity. 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. This model comparison criterion has already been applied successfully to complex models in the field of medical statistics (Zhu and Carlin 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 (Carlin and Louis 1996, Gelman, Carlin, Stern and Rubin 1996, Gilks, Richardson and Spiegelhalter 1996, Key, Pericchi and Smith 1999) 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). One 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. For specifying Bayesian stochastic volatility (SV) models, however, informative and thus *proper* prior distributions are usually employed and Bayes factors are well defined. Nonetheless, the number of unknown parameters in Bayesian SV models is large (exceeding the number of observations) because of the latent volatilities. Calculation of the Bayes factor for comparing any two models requires the marginal likelihoods and thus a marginalization over the parameter vectors in each model. If the dimension of the parameter space is large, these implicit, extremely high-dimensional integration problems pose a formidable computational challenge. In the context of SV models, Kim, Shephard and Chib (1998) and Chib, Nardari and Shephard (2002) have shown how to compute Bayes factors using the marginal likelihood approach of Chib (1995) and evaluating the marginal likelihood at the posterior mean using *particle filtering* (Kitagawa 1996, Pitt and Shephard 1999, Doucet, de Freitas and Gordon 2001). Still, it remains a computationally intensive task and 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"*.

A well-known estimate of the marginal likelihood developed by Newton and Raftery (1994) is the harmonic mean of the likelihood values. It is easy to compute and simulation-consistent but not stable because the inverse likelihood does not possess a finite variance (Chib 1995). Other shortcuts to the calculation of Bayes factors that avoid multidimensional integration through large sample approximations of $-2\ln$(Bayes-factor) include the familiar BIC, also referred to as *Schwarz Criterion* (Schwarz 1978), and the related penalized likelihood ratio model choice criterion, AIC. Either criterion requires the specification of the number of free parameters in each model. If we consider a non-hierarchical Bayesian model with parameter θ, a flat prior would correspond to a flexible and thus complex model, whereas a tight prior constrains the model. The classical definition of model complexity as the 'number of unknown parameters' could thus be considered as a special case corresponding to a non-informative prior. However, for a complex hierarchical model the specification of its dimensionality is rather arbitrary. This is typically the case for a SV model, where the parameters are augmented by the n latent volatilities with n being the sample size. As these are not independent but exhibit a Markovian dependence structure, they cannot be counted as n additional free parameters. Thus, neither BIC nor AIC are applicable for SV model comparison. As detailed in Section 3, DIC avoids this dilemma by using a complexity measure for the effective number of parameters that is based on an information theoretic argument. This quantity is readily obtained from a MCMC analysis which makes algebraic forms and large sample approximations obsolete.

Most importantly, when modelling financial time series using SV models, we agree with Box (1976) in believing that "all models are wrong but some are useful". Thus, it would be hard to specify prior model probabilities necessary for the calculation of Bayes factors. By using the DIC as a formal approach to model selection, combining a measure of fit and complexity, we can avoid this need. However, we caution in general against basing model choice solely on information criteria, as many other factors such as the model's inherent plausibility, the robustness of its inferences and model diagnostics (as for instance outlined in Kim et al. 1998, Section 4.2 and Spiegelhalter et al. 2002, Section 6) need to be taken into account. In many instances, when none of the models is clearly superior, model averaging (Hoeting, Madigan, Raftery and Volinsky 1999) might be more appropriate. Whether DIC can be used as a basis for model averaging is still an open question.

The outline of the paper is as follows: Section 2 gives an introduction to SV models, followed in Section 3 by the definition and properties of DIC. Section 4 reviews Chib (1995)'s method for calculating the marginal likelihood based on particle filtering and Newton and Raftery (1994)'s harmonic mean estimate of the marginal likelihood. In Section 5, we present results from a simulation study and compare the model ranking implied by the marginal likelihood, harmonic mean, and DIC. Section 6 applies DIC to compare the fit of various SV models to a dataset previously analyzed in the literature. We also assess the performance of DIC using the Bayes factor as a gold standard and examine the prior sensitivity of DIC. In Section 7 we present our conclusions.

2. THE STOCHASTIC VOLATILITY MODEL

In both the theoretic finance literature on option pricing and the empirical finance literature, the SV model (Hull and White 1987, Tauchen and Pitts 1983, Taylor 1982) has received much attention in recent years. It has become a powerful alternative to the ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986). Ghysel, Harvey and Renault (1996) and Shephard (1996) give an excellent review of the model.

Given a times series of daily returns $\{y_t\}_{t=1}^n$, a basic SV model consists of an observation equation

$$y_t|h_t = \exp(h_t/2)u_t, \qquad t = 1, \ldots, n, \tag{1}$$

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

$$h_t|h_{t-1} = \mu + \phi(h_{t-1} - \mu) + v_t, \qquad t = 1, \ldots, n, \tag{2}$$

which models the day-to-day variation of the volatilities as a Markov process. Here, y_t is the response variable, h_t is the log-volatility process and the errors u_t and v_t are uncorrelated Gaussian sequences with $u_t \sim N(0, 1)$ and $v_t \sim N(0, \tau^2)$. We collect the three model parameters in a vector $z = (\phi, \mu, \tau^2)$. In Section 5 and Section 6 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

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

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, Karolyi, Longstaff and Sanders 1992 and Brenner, Harjes and Kroner 1996). In other applications, for example stock market data, it is common to set this parameter equal to 0 (see Section 5 below).

Classical parameter estimation for this model is extremely difficult, because of the non-analytic form of the likelihood function. Harvey, Ruiz and Shephard (1994) and Ruiz (1994) employ a quasi-maximum likelihood technique, whereas Sandmann and Koopman (1998) use the maximum likelihood Monte Carlo method. Several method of moment approaches like the efficient method of moments (Gallant and Tauchen 1996 and Andersen, Chung and Sørensen 1999), the spectral method of moments (Knight, Satchell and Yu 2002, Singleton 2001, Viceira and Chacko 2001), the simulated method of moments (Duffie and Singleton 1993) or the 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 *ad hoc* criteria (see Andersen et al. 1999 for a review and comparison of various estimation techniques for the SV model), a Bayesian approach is based on a sound statistical paradigm. Bayesian posterior computations are performed using MCMC techniques. Several different algorithms have been proposed by Jacquier, Polson and Rossi (1994), Kim et al. (1998) and further developed in Chib et al. (2002). Although more efficient updating techniques for SV models exist, we employ the all purpose Bayesian software package BUGS based on the single-update Gibbs sampler as described in Meyer and Yu (2000) for ease of implementation. The SV model is a typical example of a hierarchical model, in which the number of unknowns, ie the parameters (z) and the unknown states ($h_1, \ldots, h_n$), exceeds the number of observations. The number of free parameters in the model could be 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 and thus precludes the use of AIC or BIC for model comparison. We will show that DIC provides an efficient and straightforward approach to defining the effective number of parameters and to identifying the most appropriate model.

3. THE DEVIANCE INFORMATION CRITERION

Assume, in general, that the distribution of the data, $y = (y_1, ..., y_n)$, depends on a p-dimensional parameter vector θ. (In the context of an SV model, θ encompasses the parameter vector z and the vector of log-volatilities $h_1, \ldots, h_n$). 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, Dempster (1974) suggested examining the posterior distribution of the classical deviance defined by

$$D(\theta) = -2 \ln f(y|\theta) + 2 \ln g(y) \tag{3}$$

for Bayesian model selection. Here, $f(y|\theta)$ is the likelihood function, ie the conditional joint probability density function of the observations given the unknown parameters, and ln $g(y)$ denotes a fully specified standardizing term that is a function of the data alone.(In our applications in Section 5 and 6, sur$g(y) = 1$.) Dempster (1974) proposed comparing plots and potential summaries such as the posterior mean of $D(\theta)$ and Spiegelhalter et al. (2002) 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:

$$\text{DIC} = \bar{D} + p_D. \tag{4}$$

1. The first term, a Bayesian measure of model fit, is defined as the posterior expectation of the deviance

$$\bar{D} = E_{\theta|y}[D(\theta)] = E_{\theta|y}[-2\ln f(y|\theta)]. \tag{5}$$

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' models.

2. The second component measures the complexity of the model by the *effective number of parameters*, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean $\bar{\theta}$ of the parameters:

$$p_D = \bar{D} - D(\bar{\theta}) = E_{\theta|y}[D(\theta)] - D(E_{\theta|y}[\theta]) = E_{\theta|y}[-2\ln f(y \mid \theta)] + 2\ln f(y|\bar{\theta}). \tag{6}$$

By defining $-2\ln f(y|\theta)$ as the residual information in the data y conditional on θ and interpreting it as a measure of uncertainty, Equation (6) shows that p_D can be regarded as the expected excess of the true over the estimated residual information in data y conditional on θ. That means we can interpret p_D as the expected reduction in uncertainty due to estimation.

By rearranging Equation (6), one gets $\bar{D} = D(\bar{\theta}) + p_D$. Thus, the DIC defined in (4) can be re-expressed as

$$\text{DIC} = D(\bar{\theta}) + 2p_D \tag{7}$$

which can be interpreted as a classical 'plug-in' measure of fit plus a measure of complexity. Therefore, the Bayesian measure of fit $\bar{D} = D(\bar{\theta}) + p_d$ already includes a penalty term for model complexity and could thus be better thought of as a measure of 'model adequacy' rather than pure goodness-of-fit.

Spiegelhalter et al. (2002) give an asymptotic justification of DIC in the case where the number of observations n grows with respect to the number of parameters p and where the prior is non-hierarchical and completely specified (ie without hyperparameters). In this situation, AIC$= D(\hat{\theta}) + 2p$, where $\hat{\theta}$ denotes the maximum likelihood (ML) estimate. This is the same formula as (7) but with the posterior mean $\bar{\theta}$ substituted by the ML estimate $\hat{\theta}$. Thus, DIC can be seen as a generalization of AIC and also compared to the Schwarz information criterion BIC$= -2 \ln f(y|\hat{\theta}) + p \ln n$. In the special case where the prior is flat, a case that corresponds to a frequentist analysis, AIC equals DIC since the ML estimate coincides with the posterior mean. In the context of normal linear regression with uncertainty in the choice of regressors, George and Foster (2000) developed empirical Bayes alternatives to penalized likelihood criteria such as AIC and BIC and Fernandez, Ley and Steel (2001) point out links of Bayes factors with classical information criteria and provide a unifying framework.

By applying a logarithmic loss function, Spiegelhalter et al. (2002) give a decision-theoretic justification for DIC and 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. (2002).

In striking contrast to calculating Bayes factors, computing DIC via MCMC is almost trivial. An estimate of $\bar{D}$ is easily calculated from the MCMC output by monitoring $D(\theta)$ and then taking the sample mean of the simulated values of $D(\theta)$. The effective number of parameters p_D can be obtained by evaluating $D(\theta)$ at the sample average of the simulated values of θ and subtracting this plug-in estimate of the deviance from the estimate of $\bar{D}$ (see also section 5.3).

So far, no efficient method has been developed for calculating reasonably accurate MC standard errors of DIC. Zhu and Carlin (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

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

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

4. MARGINAL LIKELIHOOD AND HARMONIC MEAN

Since we are going to compare the performance of DIC with that of the marginal likelihood and the harmonic mean in the next two sections, it is worthwhile to first review Chib's method for calculating the marginal likelihood and Newton and Raftery (1994)'s method for estimating the marginal likelihood by the harmonic mean of the sampled likelihood values.

4.1 Marginal Likelihood

By definition, the marginal likelihood $m(y)$ is the integral of the likelihood function with respect to the prior density $\pi(z)$

$$m(y) = \int f(y|z)\pi(z)dz \tag{8}$$

with z denoting the vector of parameters in the model. As solving this integral would require high-dimensional integration, Chib (1995) suggested evaluating the marginal likelihood by rearranging Bayes' theorem

$$m(y) = \frac{f(y|z)\pi(z)}{\pi(z|y)}$$

where $\pi(z|y)$ denotes the posterior probability density function of z. Thus, the log-marginal likelihood which is referred to as $\ln L$ in the following, can be calculated by

$$\ln L = \ln m(y) = \ln f(y|z) + \ln \pi(z) - \ln \pi(z|y), \tag{9}$$

where z is an appropriately selected high density point (in this paper we simply use the posterior mean $\bar{z}$). The first term on the right hand side of Equation (9) is the log-likelihood evaluated at the posterior mean of the parameter vector z (marginalized over the latent volatilities h_t) and the second term is the log prior density evaluated at $\bar{z}$. The third quantity involves the posterior density which is only known up to a normality constant. However, an approximation can be obtained by using a multivariate kernel density estimate as suggested in Kim et al. (1998) (see also Silverman 1986, Chapter 4) which is based on the posterior MCMC sample of z.

We mentioned in Section 2 that the log-likelihood function $\ln f(y|z)$ has no analytical form for the SV model as it is marginalized over the latent states $h_1, \ldots, h_n$, and this is why the exact maximum likelihood method is extremely difficult to implement. However, it is possible to approximate the likelihood by making use of the so-called *particle filter* method. Important contributions in this area include Gordon, Salmond and Smith (1993), Kitagawa (1996), and Pitt and Shephard (1999). By successive conditioning, the log-likelihood $\ln f(y|\bar{z})$ can be decomposed into

$$\ln f(y|\bar{z}) = \ln f(y_1|\bar{z}) + \sum_{t=1}^{n-1} \ln f(y_{t+1}|Y_t, \bar{z}) \tag{10}$$

where $Y_t = (y_1, \ldots, y_t)$ collects the available data up to time t. Taking the latent volatilities into account, each one-step-ahead prediction density has a mixture representation as

$$\begin{aligned} f(y_{t+1}|Y_t, \bar{z}) &= \int f(y_{t+1}|h_{t+1}, Y_t, \bar{z}) f(h_{t+1}|Y_t, \bar{z}) dh_{t+1} \\ &= \int f(y_{t+1}|h_{t+1}, Y_t, \bar{z}) \left[\int f(h_{t+1}|h_t, \bar{z}) f(h_t|Y_t, \bar{z}) dh_t \right] dh_{t+1} \end{aligned}$$

and can thus be estimated by

$$\frac{1}{M} \sum_{i=1}^{M} f(y_{t+1}|h_{t+1}^{(i)})$$

where $h_{t+1}^{(i)}|h_t^{(i)}$ is drawn from the state equation (2) given samples $h_t^{(i)}$, the so-called *filtered particles*, from $f(h_t|Y_t, \bar{z})$.

In this paper we utilize Kitagawa's algorithm for particle filtering which is applicable to a broad class of nonlinear non-Gaussian multi-dimensional state space models of the form,

$$\begin{cases} y_t &= H(x_t, u_t) \\ x_t &= F(x_{t-1}, v_t), \end{cases} \tag{11}$$

where x_t is a k-dimensional state vector (here, $x_t = h_t$ is the one-dimensional log-volatility), v_t is a l-dimensional white noise sequence with density $q(v)$, u_t is a one dimensional white noise sequence with density $r(u)$ and assumed uncorrelated with $\{v_s\}_{s=1}^n$, H and F are possibly nonlinear functions. Let $u_t = G(y_t, x_t)$ and G' is the derivative of G as a function of y_t. The density of the initial state vector is assumed to be $p_0(x)$. We now summarize all the steps involved in Kitagawa's algorithm:

1. Generate M l-dimensional particles from $p_0(x)$, $f_0^{(j)}$ for $j = 1, \ldots, M$.

2. Repeat the following steps for $t = 1, \ldots, n$.

   (a) Generate M l-dimensional particles from $q(v)$, $v_t^{(j)}$ for $j = 1, \ldots, M$.

   (b) Compute $p_t^{(j)} = F(f_{t-1}^{(j)}, v_t^{(j)})$ for $j = 1, \ldots, M$.

   (c) Compute $\alpha_t^{(j)} = r(G(y_t, p_t^{(j)}))$ for $j = 1, \ldots, M$.

   (d) Re-sample $\{p_t^{(j)}\}_{j=1}^M$ to get $\{f_t^{(j)}\}_{j=1}^M$ with probabilities proportional to $\{r(G(y_t, p_t^{(j)})) \times |G'(y_t, p_t^{(j)})|\}_{j=1}^M$.

It can be seen that almost all the SV models present in the next two sections can be rewritten in the state space form (11) and hence it is straightforward to modify the above algorithm to fit our need. The only exception is MODEL 5 which violates the assumption of no correlation between u_t and v_{t+1}. In the Appendix, we discuss how to modify Kitagawa's algorithm to deal with this model.

We should point out that more efficient particle filter algorithms are available. An example is the *auxiliary particle filter* introduced by Pitt and Shephard (1999). Our experience suggests that by choosing $M = 50,000$ for Kitagawa's algorithm one obtains very similar results to the auxiliary particle filter method with $M = 2,500$.

4.2 Harmonic Mean

Newton and Raftery (1994) have suggested the calculation of approximate Bayes factors for model comparison using the harmonic mean of the sampled likelihood values as a simulation consistent estimator of the required marginal likelihood. Let θ denote the parameter vector (augmented by latent volatilities), ie $\theta = (z, h_1, \ldots, h_n)$, as in Section 3. Similar to Equation (8), the marginal likelihood $m(y)$ can be expressed as

$$m(y) = \int f(y|\theta)f(\theta)d\theta$$

where $f(\theta)$ denotes the joint prior density function of θ. The importance sampling method for evaluating this integral is to generate a sample $\{\theta^{(i)}; i = 1, \ldots, M\}$ from a so-called importance sampling density $f^*(\theta)$. Under quite weak assumptions, a simulation consistent estimate of $m(y)$ is given by

$$\hat{m}(y) = \frac{\sum_{i=1}^{M} w_i f(y|\theta^{(i)})}{\sum_{i=1}^{M} w_i} \tag{12}$$

where $w_i = f(\theta^{(i)})/f^*(\theta^{(i)})$. The Gibbs sampler gives us a sample $\theta^{(i)}$ approximately drawn from the posterior density $f^*(\theta) = f(\theta|y) = \dfrac{f(y|\theta)f(\theta)}{m(y)}$. Using these $\theta^{(i)}$ in Equation (12) yields the harmonic mean estimator of $m(y)$:

$$\hat{m}_{hm}(y) = \left\{ \frac{1}{M} \sum_{i=1}^{M} \frac{1}{f(y|\theta^{(i)})} \right\}^{-1}. \tag{13}$$

$\hat{m}_{hm}(y)$ converges almost surely to the correct value $m(y)$ as M goes to infinity but it does not, in general, satisfy a Gaussian central limit theorem as $1/f(y|\theta)$ is often not square integrable with respect to the posterior distribution. Thus, a few outlying values $\theta^{(i)}$ with small likelihood values can have a large effect on the estimate.

5. A SIMULATION STUDY

The main objective of this simulation study is to see whether DIC is capable of identifying the true model from which the data are generated. Following suggestions by the referees, we also calculate the marginal likelihood and the harmonic mean estimate for each model within the set of competing models. However, we want to point out an argument by Spiegelhalter et al. (2002, rejoinder) that cautions against using the Bayes factor (or marginal likelihood) as a gold standard against which to assess DIC. The Bayes factor addresses how well the prior has predicted the observed data, whereas DIC addresses how well the posterior might predict future data generated by the same mechanism that gave rise to the observed data. Thus, these criteria cannot in general be expected to come to the same conclusions as they are designed to answer different questions. Especially for the practical selection of models of financial time series, we consider this posterior predictive outlook of DIC to be potentially more relevant.

We simulate a dataset comprising 2,000 observations from a stochastic volatility model that includes a jump component as described below. The data are plotted in the first panel of Figure 1. This SV+Jumps model (MODEL 6 in the list below) is very similar to the one proposed in the simulation analysis by Chib et al. (2002). We use the BUGS (Bayesian Inference Using Gibbs Sampling) software package (Spiegelhalter, Thomas, Best and Gilks 1996), available free of charge from

`http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

for posterior computation. 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. Although more efficient Markov Chain Monte Carlo techniques exist for fitting SV models (Kim et al. 1998), the use of BUGS is highly attractive due to the ease of implementation. In the following, we describe the list of competing models under consideration.

5.1 The Models

We fit eight different stochastic volatility models to the simulated data including the true model from which the data are generated (MODEL 6). For each of the models we list the observation and state equations (for $t = 1, \ldots, n$) and their distributional assumptions. For all cases we assume u_t and $\{v_s\}_{s=1}^n$ are uncorrelated unless we claim otherwise. Prior distributions for the unknown parameters are stated in Section 5.2.

<u>MODEL 1</u>: This model is identical to the basic SV model in Section 2.

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

with $h_0 \sim N(\mu, \tau^2)$.

<u>MODEL 2</u>: An additional non-zero mean α is added to the observation equation:

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

<u>MODEL 3</u>: An AR(2)-process for the state equation:

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

<u>MODEL 4</u>: Two independent AR(1) processes as in Gallant and Tauchen (2001) and Chernov, Gallant, Ghysel and Tauchen (2001):

$$\begin{aligned} y_t \mid h_t &= \exp(\mu/2 + h_t^{(1)}/2 + h_t^{(2)}/2)u_t, \quad u_t \overset{iid}{\sim} N(0,1), \\ h_t^{(1)} \mid h_{t-1}^{(1)}, \phi, \tau^2 &= \phi h_{t-1}^{(1)} + v_t^{(1)}, \quad v_t^{(1)} \overset{iid}{\sim} N(0, \tau^2), \\ h_t^{(2)} \mid h_{t-1}^{(2)}, \phi_2, \tau_2^2 &= \phi_2 \cdot h_{t-1}^{(2)} + v_t^{(2)}, \quad v_t^{(2)} \overset{iid}{\sim} N(0, \tau_2^2), \end{aligned}$$

Model 5: This is Model 1 including a leverage or asymmetric effect by allowing for correlation between u_t and v_{t+1}, ie

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

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) and described in Engle and Ng (1993).

Model 6: The SV+Jumps model includes a jump component and lagged observations in the observation equation:

$$\begin{aligned} y_t \mid h_t, s_t, q_t, \beta, &= \beta y_{t-1} + s_t q_t + \exp(h_t/2) u_t, \quad u_t \overset{iid}{\sim} N(0,1) \\ h_t \mid h_{t-1}, \mu, \phi, \tau^2, &= \mu + \phi(h_{t-1} - \mu) + v_t, \; v_t \overset{iid}{\sim} N(0, \tau^2) \end{aligned}$$

where q_t follows a Bernoulli distribution which takes the value of one with probability κ and zero with probability $1 - \kappa$, and $\ln(1 + s_t) \sim N(-\delta^2/2, \delta^2)$.

The underlying data are generated from this model using $\mu = -10$, $\phi = 0.96$, $\tau = 0.345$, $\beta = 0.1$, $\kappa = 0.08$, and $\delta = 0.03$.

Model 7: This model includes a jump component in the observation equation but without taking the lagged observations into consideration:

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

Model 8: Gaussian observation errors are substituted by independent central Student-t distributions with ν degrees of freedom:

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

5.2 Prior Distributions

For the parameters ϕ and τ^2 of the basic SV model, we follow exactly the prior specifications of Kim et al. (1998): $\tau^2 \sim$ Inverse-Gamma(2.5,0.025) which has a mean of 0.167 and a standard deviation of 0.024. Defining $\phi = 2\phi^* - 1$, Kim et al. (1998) specify a beta-distribution with parameters 20 and 1.5 for ϕ^* which implies a mean of 0.86 and a standard deviation of 0.11. Following Kim et al. (1998) we choose an informative but reasonably flat prior distribution for the parameter μ, a normal distribution with mean -10 and variance 25.

For α in MODEL 2 a normal distribution with mean parameter $\mu_\alpha = 0$ and variance $\sigma^2_\alpha = 10$ is specified.

For MODEL 3 we use the same prior for ϕ as for the basic SV model and center the prior for ψ around zero using a uniform distribution on [-1,1].

In MODEL 4 we again use the same prior for ϕ as for the basic SV model and center a vague prior for ϕ_2 around zero using a beta distribution with parameters 2 and 2.

The correlation parameter ρ in MODEL 5 is assumed to be uniformly distributed with support between -1 and 1.

As the parameter β in MODEL 6 is assumed to be small a priori, we use an informative normal distribution with hyperparameters $\mu_\beta = 0$ and $\sigma^2_\beta = 0.2$. The parameter q_t represents the frequency of a jump occurrence with a Bernoulli distribution with parameter κ. Following Chib et al. (2002) we specify a Beta(2,100) prior distribution which implies a mean of 0.02 and suggests that a priori on average one jump in approximately every 50th observation. Finally, as in Chib et al. (2002), we assume that $\ln(\delta)$ follows a normal prior distribution with mean -3.07 and variance 0.149.

A well-known alternative to the direct fitting of many symmetric but non-normal error distributions is through scale mixtures of normals (Andrews and Mallows 1974).

Thus, in MODEL 8 we use the alternative mixture representation of a t_d-distribution by

$$
\begin{aligned}
y_t &\sim N(0, \exp(h_t)/w_t) \\
w_t &\sim \frac{1}{\nu}\chi^2_\nu = \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})
\end{aligned}
$$

We choose a uniform distribution for ν on [2,128] as in Chib et al. (2002).

5.3 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 $(1/N)\sum_{i=1}^{N} D(\theta^{(i)})$. 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 $\bar{\theta} = (1/N)\sum_{i=1}^{N} \theta^{(i)}$. WinBUGS offers several useful convergence checking criteria available in an attached CODA (Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Best, Cowles and Vines 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). All the results we report in this paper are based on samples which have passed the Heidelberger and Welch convergence test for all parameters.

5.4 Results

In Table 1 we report means and standard errors (numbers in parentheses) of both prior and posterior distributions for each of the eight models. The numbers in square brackets are the true values of the parameters. In Table 2 we report the marginal likelihood, harmonic mean, and DIC together with $\bar{D}$ and p_D for each of the eight models as well as their associated rankings by each criterion. The results for SV MODELS 1-5 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. All calculations were performed on a Pentium-III PC, 550 MHz, running the WinBUGS 131 version updated with the DIC tool.

From the examination of these two tables we first note that the correct model (MODEL 6) is estimated by MCMC with reasonably accurate results for all six parameters. Moreover, the correct model provides the smallest value for DIC as well as for the posterior mean of the deviance despite the fact that the effective number of parameters is not the smallest. We get only a slightly larger value of DIC for the SV+Jumps model without lagged observations (MODEL 7). This is because differences between this model and the correct model are very small. Not surprisingly, this model is ranked second by DIC. All the other models perform clearly worse. For example, compared with DIC values of -14450 and -14362 for the two models with jumps, the basic SV model provides a DIC value of -13442.5. In fact, the DIC margins among all the models excluding the jump models are reasonably small. For example, DIC of the third best model differs from that of the worst model by 54.3 while the difference between the second best and the third best is 865.8. Moreover, the effective number of parameters is much larger for all the models except the jump models and none of these models fit the data as well as the jump models, indicated by the highest value for the posterior mean of the deviance. Not surprisingly the higher values of $\bar{D}$ and p_D add up to the higher DIC values.

Model 4, with two independent AR(1) components gives a relative good fit being ranked the best fitting after the jump models by DIC and the best fitting after the jump and SV-t models by marginal likelihood. It can thus be considered as a good alternative to using SV models with jumps.

Another interesting result emerging from these two tables is the performance of DIC relative to the marginal likelihood and the harmonic mean. Neither DIC nor the harmonic mean provides the same model ranking as the marginal likelihood but the differences are not substantial. Differences are not surprising as the focus of each the three criteria is different. In Chib's marginal likelihood, the focus is on a likelihood $f(y|z)$ with interest in the structure of a parameter vector z where the latent volatilities are integrated out. On the other hand, the harmonic mean estimate and DIC focus on a likelihood $f(y|\theta)$ where θ comprises both unknown parameters (z) and latent states $h_1, \ldots, h_n$. The different focus of each model corresponds to different prdiction problems of interest.

Comparison between DIC and the marginal likelihood reveals that the mixture normal-Gamma t SV model (MODEL 8) is the only cause of the discrepancy. Here it is helpful to divide DIC into a pure measure of fit $D(\bar{\theta})$ and a measure of complexity $2p_D$ as in equation (7) to see that the t SV model is heavily penalized by its high effective number of parameters. Considering $D(\bar{\theta})$, gives a value of -14715.4 for the true Model 6 and a value of -14744.1 for t SV Model 8. Thus the t SV model provides a better fit but its high complexity tips the scales. Although not reported, we have also estimated the non-scale mixture t SV model and found that the performance of these two representations are quite different. The non-scale mixture t SV model performs even worse than the scale mixture t SV model according to DIC. It has been recognized that different mixture distributions can generate different DIC values due to the fact that different mixture distributions correspond to different prediction problems and that more research and experience is needed as to the performance of DIC in the area of mixture models (Spiegelhalter et al. 2002).

Table 3 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 six runs for each of the seven models. It serves to demonstrate that DIC varies from one run to another but is reasonably stable across runs. This is in contrast to the well known instability problem of the harmonic mean.

6. AN EMPIRICAL STUDY

6.1 The Data

The dataset consists of 1512 mean-corrected daily continuously compounded returns, y_t, in decimals, on 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, Poon and Taylor (2001a) estimate the GJR-GARCH model proposed by Glosten, Jagannathan and Runkle (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 use 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, Ostdiek and Whaley (1995). Both data series are plotted in the second and third panels of Figure 1.

6.2 The Models and Prior Distributions

In this section, we fit the models introduced in section 5 to the above data set. We drop MODEL 4 from the list due to a great deal of convergence problems that we have encountered. Instead we consider as an additional extension a model that includes implied volatility:

MODEL 9: This model is very similar to the SVX model introduced in Hol and

Koopman (2000) which includes implied volatility as expressed by an additional covariate x_t:

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

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. 2001b). In the last panel of Figure 1, we plot the logarithm of absolute value of S&P100 returns which is regarded as an approximation of unobserved log-volatility. It can be seen that VIX and the logarithm of absolute value of S&P100 returns are highly correlated. Note that we demean the observations in vector x_t for convergence purposes.

A priori, λ is assumed to be uniformly distributed in the interval [-1,1]. Due to the inclusion of the implied volatility, it is not clear a priori whether the log-volatility h_t is still highly persistent. Instead of using a rather informative prior of a beta distribution with parameters 20 and 1.5 for ϕ^*, we choose a less informative prior for ϕ^*, namely, a uniform distribution with support between 0 and 1.

6.3 Results

In Table 4 we report means and standard errors (numbers in parentheses) of both prior and posterior distributions for each of the eight models. The results for Models 1-5 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. 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.

From Table 4 it can be seen that the estimated means and standard deviations for the parameters appear quite reasonable and comparable with previous estimates in the literature. For instance, in Model 1, the volatility process is estimated to be highly persistent. In Model 5 the posterior mean of ρ is -0.4139 with the upper limit of the 95% posterior credibility interval less than zero. It suggests that the leverage effect is an important feature for the S&P100 index. The parameter estimates for the two SV+Jumps models provide similar results for those parameters already covered by the SV models without jumps. As already observed in Chib et al. (2002), we note that the jump parameters κ and δ are not as precisely estimated as other parameters. However, they are well identified as their posterior distributions are substantially different from their prior distributions. The posterior mean of the jump intensity κ is 0.011 which means an average daily probability of 1.1% of a jump occurring. This implies that a jump can be expected to occur roughly every 90th day. The standard deviation of the jump size is about 0.03, i.e. daily jumps are usually around 6%.

In Model 8 the posterior mean of ν is 7.306 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. (2002) respectively. The posterior mean of λ in Model 9 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.

In Table 5 we report the marginal likelihood, harmonic mean, and DIC together with $\bar{D}$ and p_D for each of the eight models as well as their associated rankings by each criterion. The most adequate models to describe S&P100 according to DIC are the jump model without lagged observations (Model 7) and the jump model with lagged observations (Model 6), followed by the implied volatility model (Model 9) and the model including the leverage effect (Model 5). Although the posterior means of the deviance for the jump models are higher than those of most of the other models, the effective number of parameters is much lower. The effective number of parameters is around 26 for the jump models which is less than one third of the effective number of parameters for the closest competitor. MODEL 5 has the lowest posterior means of the deviance which suggests the best fit to the data. However, its effective number of parameters is much higher than for the other models. In particular, it is more than 10 times larger that of the jump models. This renders a larger value of DIC.

As for the simulated data, neither DIC nor the harmonic mean provides the same model ranking as the marginal likelihood. According to the marginal likelihood, the strength of evidence to distinguish between the models is much weaker for the S&P100 data than for the simulated data. For example, the marginal likelihood values from the second best model and the third best model only differs by 0.84 which is not worth more than a bare mention according to Jeffrey's Bayes factor scale ($\exp(0.84) = 2.316$). Nonetheless, both DIC and the marginal likelihood select MODEL 8 (ie the jump model without lagged observations) as the best performed model while the harmonic mean picks MODEL 9 (ie the t SV model).

A close look at Table 5 reveals that the mixture normal-Gamma t SV model (ie MODEL 8) is the major cause of the discrepancy between the DIC ranking and the marginal likelihood ranking. This is a similar finding to the simulated data. Another minor discrepancy arises from the first three models. The marginal likelihood ranks MODEL 2 the worst model while DIC ranks MODEL 1 the worst.

Table 6 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 six runs for each of the seven models. Again it demonstrates that DIC varies from one run to another but is reasonably stable across runs. Also, it can be seen that the ranges of DIC overlap with each other for the first three models. This explains why the first three models are difficult to distinguish.

6.4 Robustness Check

In this section we examine the implications of alternative prior distributions on DIC and the marginal likelihood. In particular, we focus on a subset of hyperparameters, namely, ϕ and κ. Also, for brevity we only consider a subset of the models, namely, the basic SV model (MODEL 1), the SV model with a leverage effect (MODEL 5), and the SV+jumps model without lagged observations (MODEL 7). Following Chib et al. (2002) we consider the following two alternative priors:

- Prior 2: $\phi^* \sim \mathrm{U}(0, 1)$;
- Prior 3: $\phi^* \sim \mathrm{U}(0, 1), \kappa \sim$ Beta with mean = 0.0385 and standard error = 0.0264.

We re-estimate all three models with Prior 2 and MODEL 7 with Prior 3 and calculate DIC and the marginal likelihood accordingly. The posterior means, standard errors, DIC, and the marginal likelihood are reported in Table 7. A comparison with the results in Table 4 shows that Prior 2 yields a posterior distribution that is almost identical to that with the original prior and that Prior 3 yields a posterior distribution that is reasonably close to that with the original prior. More importantly, DIC seems quite robust to the change of the prior. Moreover it preserves the ranking of the models considered and the ranking is consistent with that based on the marginal likelihood.

7. 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 the simulated data are quite accurate for the true model and DIC clearly identifies the correct model out of eight different alternatives. If one were to omit the mixture t SV model, DIC would give the same model ranking as the marginal likelihood. By comparing eight different SV models for the S&P100 index, comprising 1512 observations from 1993 to 1998, the jump volatility model without lagged observations turns out to be the most adequate as indicated by both DIC and the marginal likelihood. The Monte Carlo error of DIC is fairly low for all the models, thus indicating a stable performance for model comparison purposes. Finally, DIC appears robust to the change of prior distributions.

APPENDIX: PARTICLE FILTERING FOR MODEL 5

By comparing the state space model given by (11) with the specification of the SV model with a leverage effect, we have $x_t = h_t$, $H(x_t, u_t) = \exp(0.5x_t)u_t$, and $F(x_t, v_t) = \mu + \phi(x_{t-1} - \mu) + v_t$. It can be seen that the possible correlation between u_t and v_{t+1} does not impose any problem to steps 1, 2(c) and 2(d) in Kitagawa's algorithm. We now discuss how one should modify steps 2(a) and 2(b) to handle the possible correlation between u_t and v_{t+1}.

Observing that

$$\begin{pmatrix} y_{t-1} \\ v_t \end{pmatrix} \overset{iid}{\sim} N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \exp(x_{t-1}) & \rho\tau\exp(0.5x_{t-1}) \\ \rho\tau\exp(0.5x_{t-1}) & \tau^2 \end{pmatrix} \right\},$$

we have

$$v_t | y_{t-1} \sim \mathrm{N}(\rho\tau y_{t-1}\exp(-0.5x_{t-1}), \tau^2(1-\rho^2)). \tag{A.1}$$

Suppose we have a set of M realizations from $p(x_{t-1}|y_{t-1})$. We need to show how to generate a set of M realizations from $p(x_t|y_{t-1})$. Note that

$$\begin{aligned} p(x_t|y_{t-1}) &= \int\int p(x_t, x_{t-1}, v_t|y_{t-1}) dv_t dx_{t-1} \\ &= \int\int p(x_t|x_{t-1}, v_t, y_{t-1}) p(v_t|x_{t-1}, y_{t-1}) p(x_{t-1}|y_{t-1}) dv_t dx_{t-1} \\ &= \int\int p(x_t|x_{t-1}, v_t) p(v_t|y_{t-1}) p(x_{t-1}|y_{t-1}) dv_t dx_{t-1} \\ &= \int\int \delta(x_t - F(x_t, v_t)) p(v_t|y_{t-1}) p(x_{t-1}|y_{t-1}) dv_t dx_{t-1}, \end{aligned} \tag{A.2}$$

where $\delta(x)$ is the Dirac delta function. Therefore, if we generate M particles, called $v_t^{(j)}$, from density $p(v_t|y_{t-1})$ (that is a normal distribution with mean $\rho\tau y_{t-1}\exp(-0.5x_{t-1})$ and variance $\tau^2(1-\rho^2)$ according to (A.1)), we can obtain M particles by setting

$$p_t^{(j)} = F(f_{t-1}^{(j)}, v_t^{(j)}),$$

which can be regarded as independent draws from $p(x_t|y_{t-1})$.

REFERENCES

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", in *Proceedings 2nd International Symposium Information Theory*, eds. B.N. Petrov and F. Csaki, Budapest: Akademiai Kiado, pp. 267–281.

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.

Andrews, D.F. and Mallows, C.L. (1974), "Scale mixtures of normal distributions", *Journal of the Royal Statistical Society, Series B*, **36**, 99–102.

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.

Black, F. (1976), "Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Blair, B. J., Poon, S.-H. and Taylor, S. J. (2001a), "Modelling S&P 100 volatility: The information content of stock returns," *Journal of Banking and Finance*, 25, 1665–79.

Blair, B. J., Poon, S.-H. and Taylor, S. J. (2001b), "Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high frequency returns," *Journal of Econometrics*, 105, 5-26.

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

Box, G.E.P. (1976), "Science and statistics", *Journal of the American Statistical Association*, **71**, 791–799.

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**, London, Chapman and Hall.

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.

Chernov, M., Gallant, A. R., Ghysel, E., and Tauchen, G. (2001), "Alternative models for stock price dynamics", Working Paper, Department of Economics, University of North Carolina, www.unc.edu/ arg

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. (2002), "Markov Chain Monte Carlo methods stochastic volatility models," *Journal of Econometrics*, **108**, 281–316.

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.

Doucet, A., de Freitas, N. and Gordon, N. (2001), "Sequential Monte Carlo Methods in Practice", New York: Springer.

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.

Fernandez, C., Ley, E. and Steel, M.F.J. (2000), "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, **100**, 381–427.

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.

Gallant, A. R. and Tauchen, G. (2001), "Efficient method of moments", Working Paper, Department of Economics, University of North Carolina, www.unc.edu/ arg

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

George, E.I. and Foster, D.P. (2001), "Calibration and empirical Bayes variable selection," *Biometrika*, **87**, 731–747.

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

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.

Ghysels, E., Harvey, A. C., and Renault, E. (1996), "Stochastic volatility," in *Statistical Models in Finance*, eds. Rao, C.R. and Maddala, G. S., North-Holland, Amsterdam, pp. 119–191.

Gordon, N. J., Salmond, D. J. and Smith, A. E. M. (1993), "A novel approach to nonlinear and non-Gaussian Bayesian state estimation," *IEEE-Proceedings F*, **140**, 107–133.

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.

Hoeting, J.A., Madigan, A., Raftery, A.E. and Volinsky, C.T. (1999), "Bayesian model averaging: a tutorial", *Statistical Science*, **14**, 383–417.

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.

Hull, J. and White, A. (1987), "The pricing of options on asset with stochastic volatilities," *Journal of Finance*, **42**, 281–300.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), "Bayesian analysis of 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 Statistical Association*, **90**, 773–795.

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

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

Kitagawa, G. (1996), "Monte Carlo filter and smoother for Gaussian nonlinear state space models", *Journal of Computational and Graphical Statistics,* **5**, 1–25.

Knight, J.L., Satchell, S.E. and Yu, J. (2002), "Estimation of stochastic volatility model by the Empirical Characteristic Function Method," *Australian and New Zealand Journal of Statistics*, **44**, 901-917.

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.

Newton, and Raftery (1994), "Apprximate Bayesian inferences by the weighted likelihood bootstrap," *Journal of the Royal Statistical Society, Series B*, **56**, 3–48 (with discussion).

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*, London, Edward Arnold.

Pitt, M. and Shephard, N. (1999), "Filtering via simulation: Auxiliary particle filter," *The Journal of the American Statistical Association*, **94**, 590–599.

Ruiz, E. (1994), "Quasi-maximum likelihood estimation of stochastic volatility models," *Journal of Econometrics.* **63**, 289–306.

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.

Shephard, N. (1996), "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley, and O. E. Barndoff-Nielson, London, Chapman and Hall, pp. 1–67

Silverman, B. W. (1986), "Density Estimation for Statistics and Data Analysis", London, Chapman and Hall.

Singleton, K.J. (2001), "Estimation of affine asset pricing models using the empirical characteristic function," *Journal of Econometrics.* **102**, 111–141.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society, Series B*, **64**, Part 3, forthcoming.

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**, pp. 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 *Time Series Analysis: Theory and Practice, 1*, ed. Anderson, O. D., Amsterdam, North-Holland, 203–226.

Viceira, L.M. and Chacko, G. (1999), "Spectral GMM estimation of continuous-time processes," Working Paper, Harvard Business School.

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.

Table 1: Parameter Estimates for Simulated Data

		Mod 1	Mod 2	Mod 3	Mod 4	Mod 5	Mod 6	Mod 7	Mod 8
μ	Prior	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)
$[-10]$	Posterior	-9.836 (.0844)	-9.837 (.0847)	-9.845 (.0933)	-9.859 (.1615)	-9.837 (.0848)	-10.13 (.1499)	-10.08 (.1474)	-10.31 (.1512)
ϕ	Prior	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)
[.96]	Posterior	.6502 (.0398)	.6496 (.0401)	.4146 (.0703)	.9494 (.0182)	.6518 (.0399)	.9353 (.0139)	.9330 (.0147)	.9457 (.0143)
τ	Prior	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)
[.345]	Posterior	1.128 (.0617)	1.130 (.0622)	1.205 (.0567)	.2803 (.0646)	1.130 (.0624)	.3959 (.0438)	.3988 (.0450)	.3244 (.0462)
α	Prior		0.00 (3.16)						
	Posterior		3.7e-4 (.0001)						
ψ	Prior			0.00 (0.58)					
	Posterior			.2569 (.0754)					
ϕ_2	Prior				0.00 (0.45)				
	Posterior				.2135 (.0963)				
τ_2	Prior				0.38 (0.16)				
	Posterior				1.187 (.0606)				
ρ	Prior					0.00 (0.58)			
	Posterior					-0.079 (.0423)			
β	Prior						0.00 (0.45)		
[.1]	Posterior						.0807 (.0121)		
κ	Prior						0.02 (0.01)	0.02 (0.01)	
[.08]	Posterior						.0727 (.0099)	.0691 (.0100)	
δ	Prior						0.05 (0.02)	0.05 (0.02)	
[.03]	Posterior						.0342 (.0026)	.0348 (.0028)	
ν	Prior								65.0 (36.4)
	Posterior								2.563 (.1866)

Table 2: Marginal Likelihood, Harmonic Mean, DIC for Simulated Data

Model	$\ln L$		Harmonic Mean		DIC		$\bar{D}$	p_D
	Value	Ranking	Value	Ranking	Value	Ranking		
Mod 1	6472.67	7	6888.43	7	-13442.5	7	-14002.5	560.0
Mod 2	6467.51	8	6882.43	8	-13441.9	8	-14003.4	561.5
Mod 3	6474.38	5	6890.43	6	-13463.3	4	-14040.7	577.4
Mod 4	6495.51	4	6948.43	4	-13496.2	3	-14102.4	606.2
Mod 5	6472.82	6	6901.42	5	-13453.9	5	-14018.9	565.0
Mod 6	6569.16	1	7172.00	1	-14450.0	1	-14582.7	132.7
Mod 7	6548.27	2	7102.92	2	-14362.0	2	-14485.3	123.3
Mod 8	6517.44	3	6949.62	3	-13448.0	6	-14096.1	648.1

Table 3: Deviance Summaries for the Simulated Data

Model	$\overline{D}_{min}$	$\overline{D}_{max}$	p_{Dmin}	p_{Dmax}	DIC_{min}	DIC_{max}
Mod 1	-14003.5	-14000.1	557.3	560.5	-13443.8	-13441.5
Mod 2	-14004.4	-14001.1	561.5	563.4	-13443.9	-13439.7
Mod 3	-14040.7	-14036.8	574.4	577.8	-13465.4	-13461.2
Mod 4	-14103.9	-14100.2	603.3	606.4	-13499.7	-13495.6
Mod 5	-14022.9	-14018.9	565.0	568.1	-13456.5	-13453.9
Mod 6	-14585.7	-14580.0	131.7	133.4	-14452.0	-14448.6
Mod 7	-14485.3	-14479.5	123.3	124.7	-14362.0	-14354.8
Mod 8	-14099.3	-14096.0	648.1	650.2	-13453.1	-13448.1

Table 4: Parameter Estimates for S&P100 Data

		Mod 1	Mod 2	Mod 3	Mod 5	Mod 6	Mod 7	Mod 8	Mod 9
μ	Prior	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)	-10.0 (5.00)
	Posterior	-9.971 (.2573)	-9.956 (.2408)	-9.986 (.2663)	-9.951 (.2019)	-10.06 (.2889)	-10.07 (.2897)	-10.34 (.3213)	-9.942 (.0460)
ϕ	Prior	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.86 (0.11)	0.00 (0.58)
	Posterior	.9803 (.0088)	.9789 (.0094)	.8375 (.1525)	.9743 (.0100)	.9868 (.0070)	.9873 (.0069)	.9923 (.0044)	-.2745 (.1155)
τ	Prior	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)	0.12 (0.05)
	Posterior	.1674 (.0297)	.1729 (.0317)	.1886 (.0442)	.1947 (.0319)	.1331 (.0284)	.1302 (.0286)	.1005 (.0189)	0.4375 (.0763)
α	Prior		0.00 (3.16)						
	Posterior		1.03e-4 (1.7e-4)						
ψ	Prior			0.00 (0.58)					
	Posterior			.1413 (.1495)					
ρ	Prior				0.00 (0.58)				
	Posterior				-.4139 (.0860)				
β	Prior					0.00 (0.45)			
	Posterior					.0050 (.0263)			
κ	Prior					0.02 (0.01)	0.02 (0.01)		
	Posterior					.0114 (.0076)	.0115 (.0075)		
δ	Prior					0.05 (0.02)	0.05 (0.02)		
	Posterior					.0315 (.0119)	.0324 (.0121)		
ν	Prior							65.0 (36.4)	
	Posterior							7.306 (1.532)	
λ	Prior								0.0 (0.58)
	Posterior								.1527 (.0159)

Table 5: Marginal Likelihood, Harmonic Mean, DIC for S&P100 Data

Model	$\ln L$		Harmonic Mean		DIC		$\bar{D}$	p_D
	Value	Ranking	Value	Ranking	Value	Ranking		
Mod 1	5227.45	6	5270.40	8	-10529.6	8	-10613.1	83.5
Mod 2	5223.17	8	5276.64	7	-10530.0	7	-10616.9	86.9
Mod 3	5226.17	7	5277.29	6	-10531.5	6	-10616.7	85.2
Mod 5	5229.87	5	5300.43	3	-10610.9	4	-10936.3	325.4
Mod 6	5241.54	3	5279.58	5	-10649.7	2	-10677.2	27.5
Mod 7	5243.76	1	5288.27	4	-10653.5	1	-10679.3	25.8
Mod 8	5242.38	2	5331.00	1	-10566.1	5	-10788.9	222.8
Mod 9	5234.48	4	5301.30	2	-10618.8	3	-10741.3	122.5

Table 6: Deviance Summaries for the S&P100 Data

Model	$\overline{D}_{min}$	$\overline{D}_{max}$	p_{Dmin}	p_{Dmax}	DIC_{min}	DIC_{max}
Mod 1	-10617.1	-10611.6	81.9	85.8	-10531.3	-10527.4
Mod 2	-10618.4	-10615.9	85.1	88.2	-10531.4	-10529.6
Mod 3	-10621.3	-10613.1	83.1	88.8	-10532.8	-10530.0
Mod 5	-10941.4	-10934.9	323.0	328.6	-10613.3	-10610.9
Mod 6	-10681.5	-10674.9	24.9	29.8	-10656.7	-10645.1
Mod 7	-10680.7	-10675.9	24.1	30.3	-10655.1	-10645.7
Mod 8	-10791.5	-10787.3	222.2	226.1	-10565.5	-10566.9
Mod 9	-10741.5	-10738.8	120.8	124.0	-10618.8	-10617.1

Table 7: Sensitivity of DIC and Marginal Likelihood to the Prior

	Model 1	Model 5	Model 7	
	Prior 2	Prior 2	Prior 2	Prior 3
μ	-9.970 (0.2543)	-9.963 (0.2240)	-10.10 (0.3250)	-10.20 (0.3232)
ϕ	0.9806 (0.0092)	0.9768 (0.0098)	0.9886 (0.00678)	0.9887 (0.00694)
τ	0.1680 (0.0327)	0.1865 (0.0317)	0.1271 (0.0271)	0.1266 (0.0290)
ρ		-0.4145 (0.0883)		
κ			0.0107 (0.0064)	0.0121 (0.0080))
δ			0.0337 (0.0113)	0.0298 (0.0120)
DIC	-10530.7	-10618.1	-10646.5	-10659.0
$\ln L$	5225.39	5228.03	5242.27	5240.88

Figure 1: Time Series Plots for Simulated Data, S&P100, VIX, Logarithm of Absolute Value of S&P100 Returns