Australian National University

Crawford School of Public Policy

CAMA

Centre for Applied Macroeconomic Analysis

Issues in Comparing Stochastic Volatility Models Using the Deviance Information Criterion

CAMA Working Paper 51/2014
July 2014

Joshua C.C. Chan
Research School of Economics, ANU and
Centre for Applied Macroeconomic Analysis (CAMA), ANU

Angelia L. Grant
Centre for Applied Macroeconomic Analysis (CAMA), ANU

Abstract

The deviance information criterion (DIC) has been widely used for Bayesian model comparison. In particular, a popular metric for comparing stochastic volatility models is the DIC based on the conditional likelihood—obtained by conditioning on the latent variables. However, some recent studies have argued against the use of the conditional DIC on both theoretical and practical grounds. We show via a Monte Carlo study that the conditional DIC tends to favor overfitted models, whereas the DIC calculated using the observed-data likelihood—obtained by integrating out the latent variables—seems to perform well. The main challenge for obtaining the latter DIC for stochastic volatility models is that the observed-data likelihoods are not available in closed-form. To overcome this difficulty, we propose fast algorithms for estimating the observed-data likelihoods for a variety of stochastic volatility models using importance sampling. We demonstrate the methodology with an application involving daily returns on the Standard & Poors (S&P) 500 index.

Keywords

Bayesian model comparison, nonlinear state space, DIC, jumps, moving average, S&P 500

JEL Classification

C11, C15, C52, C58

Address for correspondence:

(E) cama.admin@anu.edu.au

Issues in Comparing Stochastic Volatility Models Using the Deviance Information Criterion

Joshua C.C. Chan
Research School of Economics
and Centre for Applied Macroeconomic Analysis,
Australian National University

Angelia L. Grant
Centre for Applied Macroeconomic Analysis,
Australian National University

July 2014

Abstract

The deviance information criterion (DIC) has been widely used for Bayesian model comparison. In particular, a popular metric for comparing stochastic volatility models is the DIC based on the conditional likelihood—obtained by conditioning on the latent variables. However, some recent studies have argued against the use of the conditional DIC on both theoretical and practical grounds. We show via a Monte Carlo study that the conditional DIC tends to favor overfitted models, whereas the DIC calculated using the observed-data likelihood—obtained by integrating out the latent variables—seems to perform well. The main challenge for obtaining the latter DIC for stochastic volatility models is that the observed-data likelihoods are not available in closed-form. To overcome this difficulty, we propose fast algorithms for estimating the observed-data likelihoods for a variety of stochastic volatility models using importance sampling. We demonstrate the methodology with an application involving daily returns on the Standard & Poors (S&P) 500 index.

Keywords: Bayesian model comparison, nonlinear state space, DIC, jumps, moving average, S&P 500

JEL classification: C11, C15, C52, C58

1 Introduction

Stochastic volatility models are widely used for modeling financial time series, and have more recently become important in macroeconometric modeling following the seminal work of Cogley and Sargent (2005) and Primiceri (2005). As a result, there is now a large and growing family of flexible stochastic volatility models (see, e.g., Chib, Nardari, and Shephard, 2002; Koopman and Hol Uspensky, 2002; Jensen and Maheu, 2010; Nakajima and Omori, 2012; Chan, 2013; Mumtaz and Zanetti, 2013; Eisenstat and Strachan, 2014, to name but a few examples). Given the wide range of model candidates, it has become increasingly important to be able to discriminate between competing models for a given application.

One popular metric for Bayesian model comparison is the deviance information criterion (DIC) proposed by Spiegelhalter, Best, Carlin, and van der Linde (2002). Celeux, Forbes, Robert, and Titterington (2006) later point out that there are numerous alternative definitions of the DIC depending on different concepts of the likelihood. In particular, the DIC based on the conditional likelihood—obtained by conditioning on the latent variables—has been widely used for comparing stochastic volatility models, following the influential paper by Berg, Meyer, and Yu (2004). The popularity of the conditional DIC in this setting is partly due to its trivial computation and its implementation in standard software packages, including WinBUGS. It has been used to compare a wide variety of stochastic volatility models in empirical applications; recent studies include Yu and Meyer (2006), Abanto-Valle, Bandyopadhyay, Lachos, and Enriquez (2010), Vo (2011), Mumtaz and Surico (2012), Tsiotas (2012), Brooks and Prokopczuk (2013) and Wang, Choy, and Chan (2013).

Despite its popularity, some recent papers have criticized the use of the conditional DIC on both theoretical and practical grounds. Li, Zeng, and Yu (2012) argue that the conditional DIC should not be used as a model selection criterion, as the conditional likelihood of the augmented data is nonregular and hence invalidates the standard asymptotic arguments that are needed to justify the DIC. On practical grounds, Millar (2009) provides a Monte Carlo study in which various Poisson models are compared, and the conditional DIC almost always favors the Poisson-gamma model instead of the Poisson-lognormal model, even when data are simulated from the latter. In another study, Chan and Grant (2014) show—through empirical examples involving macroeconomic and financial data—that the conditional DICs typically have large numerical standard errors. Hence, it might not be computationally feasible to estimate the conditional DIC accurately enough to be useful as a model comparison criterion. In light of this finding, they recommend that numerical standard errors of the DIC estimates should be reported, which is rarely done in empirical research.

We contribute to this line of research by showing via a Monte Carlo study that the conditional DIC tends to prefer overfitted stochastic volatility models. This is an important finding given that the conditional DIC is widely used in empirical applications. In contrast, the DIC based on the observed-data likelihood—obtained by integrating out the

latent variables—seems to be able to select the correct model. This result is not surprising as standard asymptotic arguments for justifying the DIC apply to the observed-data DIC. The main challenge for obtaining the latter DIC for stochastic volatility models is that the observed-data likelihoods are not available in closed-form, which is the main reason why it is not used in practice. We take a first step to address this difficulty by proposing efficient algorithms for estimating the observed-data likelihoods for a variety of stochastic volatility models using importance sampling. A key feature of our approach is that it draws on recent advances in band matrix algorithms rather than using the conventional Kalman filter. By carefully constructing a good importance sampling estimator for the observed-data likelihood, we show that the observed-data DIC can be accurately estimated. Furthermore, these observed-data likelihood estimators can be used in other settings, such as in developing more efficient Markov chain Monte Carlo (MCMC) samplers for estimation and algorithms for computing the marginal likelihood.

The rest of this paper is organized as follows. Section 2 outlines the concept of deviance and two definitions of the DIC. In Section 3 we discuss various stochastic volatility models that are widely used in the literature and their estimation. In Section 4 we propose importance sampling algorithms—based on fast matrix routines—for estimating the observed-data likelihoods for stochastic volatility models. The proposed methods are demonstrated via a Monte Carlo study in Section 5. Moreover, the behavior of the conditional and observed-data DICs are also examined. Section 6 illustrates the methodology with an application involving daily returns on the S&P 500. Directions for future research are briefly discussed in Section 7.

2 Deviance Information Criterion

The seminal paper by Spiegelhalter, Best, Carlin, and van der Linde (2002) introduces and develops the concept of deviance information criterion (DIC) for model comparison. This criterion is based on the *deviance*, which is defined as

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} \,|\, \boldsymbol{\theta}) + 2 \log h(\mathbf{y}),$$

where $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ is the likelihood function of the parametric model and $h(\mathbf{y})$ is some fully specified standardizing term that is a function of the data alone. The *effective number of parameters* p_D of the parametric model is defined to be

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\widetilde{\boldsymbol{\theta}}),$$

where

$$\overline{D(\boldsymbol{\theta})} = -2\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}) \,|\, \mathbf{y}] + 2 \log h(\mathbf{y})$$

is the posterior mean deviance and $\widetilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$, which is typically taken as the posterior mean or mode. Then, the *deviance information criterion* is defined as the sum of the posterior mean deviance, which can be used as a Bayesian measure of model fit or

adequacy, and the effective number of parameters that measures model complexity:

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D.$$

Hence, the DIC may be viewed as a trade-off between model adequacy and complexity. For model comparison, the function $h(\mathbf{y})$ is often set to be unity for all models. Therefore, the DIC becomes

$$\text{DIC} = -4\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}) \,|\, \mathbf{y}] + 2 \log f(\mathbf{y} \,|\, \widetilde{\boldsymbol{\theta}}).$$

Given a set of competing models for the data, the preferred model is the one with the minimum DIC value.

For latent variable models, such as stochastic volatility models, Celeux, Forbes, Robert, and Titterington (2006) point out that there are numerous alternative definitions of the DIC depending on different concepts of the likelihood. In particular, suppose we augment the model $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ with a vector of latent variables $\mathbf{z}$ with density $f(\mathbf{z} \,|\, \boldsymbol{\theta})$ such that

$$f(\mathbf{y} \,|\, \boldsymbol{\theta}) = \int f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \,|\, \boldsymbol{\theta}) \mathrm{d}\mathbf{z} = \int f(\mathbf{y}, \mathbf{z} \,|\, \boldsymbol{\theta}) \mathrm{d}\mathbf{z},$$

where $f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{z})$ is the *conditional likelihood* and $f(\mathbf{y}, \mathbf{z} \,|\, \boldsymbol{\theta})$ is the *complete-data likelihood*. We refer to the likelihood $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ as the *observed-data likelihood* or the *integrated likelihood*.

Naturally, one can define the DIC using the observed-data likelihood and we call this the observed-data DIC:

$$\text{DIC}_{\text{obs}} = -4\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}) \,|\, \mathbf{y}] + 2 \log f(\mathbf{y} \,|\, \widehat{\boldsymbol{\theta}}), \tag{1}$$

where the estimate $\widetilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is set as the posterior mode $\widehat{\boldsymbol{\theta}}$. The term $\mathbb{E}_{\boldsymbol{\theta}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}) \,|\, \mathbf{y}]$ can be estimated by averaging the log-observed-data likelihoods $\log f(\mathbf{y} \,|\, \boldsymbol{\theta})$ over the posterior draws of $\boldsymbol{\theta}$. In addition, the posterior mode $\widehat{\boldsymbol{\theta}}$ is often approximated by the draw that has the highest value of $f(\mathbf{y} \,|\, \boldsymbol{\theta}) f(\boldsymbol{\theta})$ among the posterior draws, where $f(\boldsymbol{\theta})$ is the prior density.

One main difficulty in computing DIC_{obs} is that the observed-data likelihood $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ is typically time-consuming to evaluate for a wide variety of latent variable models (although important exceptions exist, see, e.g., Chan and Grant, 2014). Since the latent variable structure is usually chosen so that the conditional likelihood $f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{z})$ is available in closed-form, one can alternatively define the DIC using the conditional likelihood and we refer to this version as the conditional DIC:

$$\text{DIC}_{\text{con}} = -4\mathbb{E}_{\boldsymbol{\theta},\mathbf{z}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{Z}) \,|\, \mathbf{y}] + 2 \log f(\mathbf{y} \,|\, \widehat{\mathbf{z}}, \widehat{\boldsymbol{\theta}}), \tag{2}$$

where $(\widehat{\mathbf{z}}, \widehat{\boldsymbol{\theta}})$ is the joint maximum a posteriori (MAP) estimate of the pair $(\mathbf{z}, \boldsymbol{\theta})$ given the data $\mathbf{y}$. As before, the expectation $\mathbb{E}_{\boldsymbol{\theta},\mathbf{z}}[\log f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{Z}) \,|\, \mathbf{y}]$ can be estimated by averaging the log-conditional likelihoods $\log f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{z})$ over the posterior draws of the pair $(\mathbf{z}, \boldsymbol{\theta})$. Moreover, the joint MAP estimate can be approximated by the best pair among the posterior draws, i.e., the pair that has the highest value of $f(\mathbf{y}, \mathbf{z} \,|\, \boldsymbol{\theta}) f(\boldsymbol{\theta})$.

Following the influential paper by Berg, Meyer, and Yu (2004), the conditional DIC is widely used for comparing stochastic volatility models, whereas the observed-data DIC is not computed in practice due to the difficulty in evaluating the observed-data likelihood. However, despite its popularity, in Section 5 we show via a Monte Carlo study that the conditional DIC tends to favor overfitted models. In contrast, the observed-data DIC seems to perform well and is better able to choose the correct model.

3 Stochastic Volatility Models

In this section, we first discuss various stochastic volatility models that are widely used in the literature for modeling financial and macroeconomic time series. Then we outline some efficient algorithms for fitting these models that build on fast band matrix routines.

3.1 The Models

We consider five different stochastic volatility models. The first model is the standard stochastic volatility model, which we denote as SV:

$$y_t = \mu + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{3}$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \omega_h^2). \tag{4}$$

The log-volatility h_t follows a stationary AR(1) process with $|\phi_h| < 1$ and is initialized with $h_1 \sim \mathcal{N}(\mu_h, \omega_h^2/(1-\phi_h^2))$.

Under the second model considered, which we refer to as SV2, the observation equation is the same as in (3), but instead of the log-volatility h_t following an AR(1) process as in (4), it follows a stationary AR(2) process:

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \rho_h(h_{t-2} - \mu_h) + \varepsilon_t^h, \qquad \varepsilon_t^h \sim \mathcal{N}(0, \omega_h^2), \tag{5}$$

where we assume the roots of the characteristic polynomial associated with (ϕ_h, ρ_h) lie outside the unit circle. Further, the process is initialized by

$$h_1, h_2 \sim \mathcal{N}\left(\mu_h, \frac{(1-\rho_h)\omega_h^2}{(1+\rho_h)((1-\rho_h)^2 - \phi_h^2)}\right).$$

The third model considered allows for the possibility of infrequent "jumps" in the data series, which may be important for high frequency financial data. Under the stochastic volatility model with jumps (SVJ), the observation equation becomes:

$$y_t = \mu + k_t q_t + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{6}$$

where q_t is a Bernoulli jump random variable with success probability $\mathbb{P}(q_t = 1) = \kappa$ and the jump size k_t is modeled as $\log(1 + k_t) \sim \mathcal{N}(-0.5\delta^2, \delta^2)$ so that its expectation is zero. The log-volatility h_t follows the same AR(1) process as in (4).

Another variant is the stochastic volatility in mean (SVM) model of Koopman and Hol Uspensky (2002), which is often used to study volatility feedback. Specifically, under the SVM model, the stochastic volatility enters the observation equation as a covariate:

$$y_t = \mu + \alpha \mathrm{e}^{h_t} + \varepsilon_t^y, \qquad \varepsilon_t^y \sim \mathcal{N}(0, \mathrm{e}^{h_t}). \tag{7}$$

As before, the log-volatility follows the same AR(1) process as in (4).

The final model considered is a version of the stochastic volatility models with moving average errors in Chan (2013). Specifically, consider the following first-order moving average model with stochastic volatility:

$$y_t = \mu + \varepsilon_t^y, \tag{8}$$

$$\varepsilon_t^y = u_t + \psi u_{t-1}, \qquad u_t \sim \mathcal{N}(0, \mathrm{e}^{h_t}), \tag{9}$$

where we assume that $u_0 = 0$ and the invertibility condition is satisfied, i.e., $|\psi| < 1$. Again the log-volatility h_t is assumed to follow the AR(1) process as in (4). This stochastic volatility model is referred to as SVMA. We summarize the five stochastic volatility models in Table 1.

Table 1: List of stochastic volatility models.

Model	Description
SV	standard stochastic volatility model where h_t follows a stationary AR(1)
SV2	same as SV but h_t follows a stationary AR(2)
SVJ	same as SV but the observation equation contains a "jump" component
SVM	same as SV but h_t enters the observation equation as a covariate
SVMA	same as SV but the observation error follows an MA(1)

We now discuss the set of priors considered under each of the models. For the standard SV, we assume the following independent priors for μ, μ_h, ϕ_h and ω_h^2:

$$\begin{aligned} \mu &\sim \mathcal{N}(\mu_0, V_\mu), & \mu_h &\sim \mathcal{N}(\mu_{h0}, V_{\mu_h}), \\ \phi_h &\sim \mathcal{N}(\phi_{h0}, V_{\phi_h}) \mathbb{1}(|\phi_h| < 1), & \omega_h^2 &\sim \mathcal{IG}(\nu_h, S_h), \end{aligned} \tag{10}$$

where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse-gamma distribution. Note that we impose the stationarity condition $|\phi_h| < 1$ through the prior on ϕ_h. For the SV2, we use the same priors for μ, μ_h and ω_h^2 as in (10), but replace the prior for ϕ_h with a prior for $\boldsymbol{\theta}_h = (\phi_h, \rho_h)'$: $\boldsymbol{\theta}_h \sim \mathcal{N}(\boldsymbol{\theta}_{h0}, \mathbf{V}_{\boldsymbol{\theta}_h}) \mathbb{1}(\boldsymbol{\theta}_h \in \mathbf{A})$, where $\mathbf{A} \subset \mathbb{R}^2$ is the set where the roots of the characteristic polynomial defined by $\boldsymbol{\theta}_h$ lie outside the unit circle.

For each of the remaining models, the priors for μ, μ_h, ϕ_h and ω_h^2 are exactly the same as in (10). In addition, under the SVJ, the jump intensity κ is assumed to have a beta distribution and the jump variance δ follows a log-normal distribution: $\kappa \sim \mathcal{B}(k_a, k_b)$ and $\log \delta \sim \mathcal{N}(\delta_0, V_\delta)$. For the SVM, the coefficient of the volatility is assumed to have a normal distribution: $\alpha \sim \mathcal{N}(\alpha_0, V_\alpha)$. Finally, the MA(1) coefficient in the SVMA has a normal distribution truncated within the unit interval: $\psi \sim \mathcal{N}(\psi_0, V_\psi) \mathbb{1}(|\psi| < 1)$.

3.2 Bayesian Estimation

In this section, we discuss a general approach for fitting all the stochastic volatility models in Section 3.1. The main difficulty in the estimation is the step where one simulates from the joint distribution of $\mathbf{h} = (h_1, \ldots, h_T)'$ conditional on the data and other model parameters, as the observation equation is nonlinear in $\mathbf{h}$. A key feature of our approach is that it builds upon fast band and sparse matrix algorithms rather than using the conventional Kalman filter. Recent papers using the former approach include Rue (2001) for linear Gaussian Markov random fields; Chan and Jeliazkov (2009) and McCausland, Miller, and Pelletier (2011) for linear Gaussian state space models; Rue, Martino, and Chopin (2009) for nonlinear Markov random fields; and McCausland (2012), Chan, Koop, and Potter (2013), Chan (2014) and Djegnéné and McCausland (2014) for nonlinear state space models.

More specifically, our approach exploits the special structure of the problem, namely, that the Hessian of the log-conditional density of $\mathbf{h}$ is a band matrix—i.e., it contains only a few nonzero elements along a narrow diagonal band. This feature is important in developing efficient sampling algorithms. In addition, the same approach can be used for obtaining efficient importance sampling estimators as discussed in Section 4. For concreteness, we focus on the estimation of the standard stochastic volatility model in (3)–(4), with modifications of the main algorithm for fitting the other models discussed in Appendix A. Let $\mathbf{y} = (y_1, \ldots, y_T)'$. Then posterior draws can be obtained by sequentially sampling from:

1. $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$;
2. $p(\mu \,|\, \mathbf{y}, \mathbf{h}, \mu_h, \phi_h, \omega_h^2) = p(\mu \,|\, \mathbf{y}, \mathbf{h})$;
3. $p(\mu_h \,|\, \mathbf{y}, \mu, \mathbf{h}, \phi_h, \omega_h^2) = p(\mu_h \,|\, \mathbf{h}, \phi_h, \omega_h^2)$;
4. $p(\omega_h^2 \,|\, \mathbf{y}, \mu, \mathbf{h}, \mu_h, \phi_h) = p(\omega_h^2 \,|\, \mathbf{h}, \mu_h, \phi_h)$;
5. $p(\phi_h \,|\, \mathbf{y}, \mu, \mathbf{h}, \mu_h, \omega_h^2) = p(\phi_h \,|\, \mathbf{h}, \mu_h, \omega_h^2)$.

In Step 1, the joint conditional density $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ is high-dimensional and non-standard. For the standard stochastic volatility model, this step can be accomplished using the auxiliary mixture sampler of Kim, Shepherd, and Chib (1998). However, this approach is model specific and cannot be easily generalized to estimate other stochastic volatility models such as the SVM. Here we discuss a direct method to simulate from this density using the acceptance-rejection Metropolis-Hastings algorithm (see, e.g., Tierney, 1994). More specifically, we note that the Hessian of $\log p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ is a band matrix. Consequently, using fast band matrix routines we can quickly obtain a Gaussian approximation as a proposal density. Furthermore, by construction, the precision matrix—i.e., the inverse of the covariance matrix—of the Gaussian proposal density is also a band matrix. As such, candidate draws can be obtained quickly via the precision

sampler in Chan and Jeliazkov (2009) instead of Kalman filter-based algorithms. The computation details are given in Appendix A.

Steps 2, 3 and 4 can be easily completed, as all the conditional distributions are standard. In particular, it is easy to check that

$$(\mu \,|\, \mathbf{y}, \mathbf{h}) \sim \mathcal{N}(\widehat{\mu}, D_\mu), \quad (\mu_h \,|\, \mathbf{h}, \phi_h, \omega_h^2) \sim \mathcal{N}(\widehat{\mu}_h, D_{\mu_h}), \quad (\omega_h^2 \,|\, \mathbf{h}, \mu_h, \phi_h) \sim \mathcal{IG}(\nu_h + T/2, \widetilde{S}_h),$$

where $\widetilde{S}_h = S_h + ((1-\phi_h^2)(h_1-\mu_h)^2 + \sum_{t=2}^T (h_t - \mu_h - \phi_h(h_{t-1}-\mu_h))^2)/2$,

$$D_\mu^{-1} = V_\mu^{-1} + \sum_{t=1}^T \mathrm{e}^{-h_t}, \quad \widehat{\mu} = D_\mu(V_\mu^{-1}\mu_0 + \sum_{t=1}^T \mathrm{e}^{-h_t} y_t),$$
$$D_{\mu_h}^{-1} = V_{\mu_h}^{-1} + \mathbf{X}_{\mu_h}' \mathbf{\Sigma}_\mathbf{h}^{-1} \mathbf{X}_{\mu_h}, \quad \widehat{\mu}_h = D_{\mu_h}(V_{\mu_h}^{-1}\mu_{h0} + \mathbf{X}_{\mu_h}' \mathbf{\Sigma}_\mathbf{h}^{-1} \mathbf{z}_{\mu_h}),$$

with $\mathbf{X}_{\mu_h} = (1, 1-\phi_h, \ldots, 1-\phi_h)'$, $\mathbf{z}_{\mu_h} = (h_1, h_2 - \phi_h h_1, \ldots, h_T - \phi_h h_{T-1})'$ and $\mathbf{\Sigma}_\mathbf{h} = \mathrm{diag}(\omega_h^2/(1-\phi_h^2), \omega_h^2, \ldots, \omega_h^2)$.

Lastly, one can sample from $p(\phi_h \,|\, \mathbf{h}, \mu_h, \omega_h^2)$ using an independence-chain Metropolis-Hastings step with proposal $\mathcal{N}(\widehat{\phi}_h, D_{\phi_h})\mathbb{1}(|\phi_h| < 1)$, where $D_{\phi_h}^{-1} = V_{\phi_h}^{-1} + \mathbf{X}_{\phi_h}' \mathbf{X}_{\phi_h}/\omega_h^2$ and $\widehat{\phi}_h = D_{\phi_h}(V_{\phi_h}^{-1}\phi_{h0} + \mathbf{X}_{\phi_h}' \mathbf{z}_{\phi_h}/\omega_h^2)$, with $\mathbf{X}_{\phi_h} = (h_1 - \mu_h, \ldots, h_{T-1} - \mu_h)'$ and $\mathbf{z}_{\phi_h} = (h_2 - \mu_h, \ldots, h_T - \mu_h)'$.

4 Importance Sampling for the Observed-Data Likelihoods

The popularity of the conditional DIC for comparing stochastic volatility models is partly due to its straightforward computation and its implementation in standard software such as WinBUGS. On the other hand, computing the observed-data DIC is less straightforward. In a recent paper, Chan and Grant (2014) derive analytical expressions for the observed-data likelihoods for a variety of linear latent variable models. However, for the stochastic volatility models discussed in Section 3, the observed-data likelihoods are not available in closed-form. One option, at least in principle, is the auxiliary particle filter proposed in Pitt and Shephard (1999), which can be used to evaluate the observed-data likelihood for general nonlinear state space models. In practice, however, the auxiliary particle filter is computationally intensive and it might not be feasible to be employed in our setting as the observed-data likelihood needs to be evaluated tens of thousands of times. To overcome this difficulty, in this section we consider fast algorithms for estimating the observed-data likelihoods for stochastic volatility models using importance sampling (see, e.g., Kroese, Taimre, and Botev, 2011, Chapter 9.7).

Recall that the observed-data or integrated likelihood is given by

$$f(\mathbf{y} \,|\, \boldsymbol{\theta}) = \int f(\mathbf{y} \,|\, \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \,|\, \boldsymbol{\theta}) \mathrm{d}\mathbf{z},$$

where $f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z})$ is the conditional likelihood and $f(\mathbf{z} \mid \boldsymbol{\theta})$ is the prior density of the latent variables $\mathbf{z}$. Let $g(\mathbf{z})$ be a density that dominates $f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \mid \boldsymbol{\theta})$, i.e., $g(\mathbf{z}) = 0$ implies $f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \mid \boldsymbol{\theta}) = 0$. Then, the observed-data likelihood can be rewritten as

$$f(\mathbf{y} \mid \boldsymbol{\theta}) = \int \frac{f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \mid \boldsymbol{\theta})}{g(\mathbf{z})} g(\mathbf{z}) \mathrm{d}\mathbf{z}.$$

Hence, if $\mathbf{Z}_1, \ldots, \mathbf{Z}_R$ are independent samples from the *importance density* $g(\mathbf{z})$, then

$$\widehat{f(\mathbf{y} \mid \boldsymbol{\theta})} = \frac{1}{R} \sum_{i=1}^{R} \frac{f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{Z}_i) f(\mathbf{Z}_i \mid \boldsymbol{\theta})}{g(\mathbf{Z}_i)} \tag{11}$$

is an unbiased, simulation-consistent estimator of the observed-data likelihood $f(\mathbf{y} \mid \boldsymbol{\theta})$. The quality of the importance sampling estimator in (11) depends critically on the choice of the importance density $g(\mathbf{z})$. It can be shown that the conditional density of the latent variables $f(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}) \propto f(\mathbf{y} \mid \boldsymbol{\theta}, \mathbf{z}) f(\mathbf{z} \mid \boldsymbol{\theta})$ gives rise to a zero-variance importance sampling estimator (see, e.g., Kroese, Taimre, and Botev, 2011, Chapter 9.7.1). An obvious difficulty is that the evaluation of the optimal importance density $f(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta})$ is not possible for stochastic volatility models as the normalization constant is not known. However, it provides guidance for choosing a good importance density. In particular, we would like to choose $g(\mathbf{z})$ to be "close" to the optimal importance density $f(\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta})$. In what follows, we focus on the standard stochastic volatility model, with the importance densities for the other stochastic volatility models discussed in Appendix B.

For the standard stochastic volatility model in (3)–(4), the latent variables are the log-volatilities $\mathbf{h}$. Therefore, we wish to approximate the conditional density

$$p(\mathbf{h} \mid \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) \propto f(\mathbf{y} \mid \mu, \mathbf{h}) f(\mathbf{h} \mid \mu_h, \phi_h, \omega_h^2)$$

to obtain a good importance density $g(\mathbf{h})$ for the estimator in (11). In fact, we have already discussed such an approximation when we outlined the estimation of the stochastic volatility model in Section 3.2. Specifically, we considered (for details see Appendix A) the Gaussian approximation with mean vector $\widehat{\mathbf{h}}$ and precision matrix $\mathbf{K}_\mathbf{h}$, where $\widehat{\mathbf{h}}$ is the mode of $p(\mathbf{h} \mid \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ and $\mathbf{K}_\mathbf{h}$ is the negative Hessian evaluated at the mode. Note that this approximating Gaussian density is the same as the one proposed in Durbin and Koopman (1997), although we obtain the approximation via band matrix routines instead of the Kalman filter.

In addition, note that $\mathbf{K}_\mathbf{h}$ is a band matrix. As such, draws from $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_\mathbf{h}^{-1})$ can be efficiently obtained using the precision sampler in Chan and Jeliazkov (2009), where the computation cost of obtaining an additional draw is only $\mathcal{O}(T)$. This is a crucial feature as multiple draws from the high-dimensional importance density are required to construct the estimator in (11). In addition, this importance density can be quickly evaluated at any point as its precision matrix $\mathbf{K}_\mathbf{h}$ is a band matrix. Choices of importance densities for the other stochastic volatility models are discussed in Appendix B.

5 A Monte Carlo Study

In this section, we examine the behavior of the conditional and observed-data DICs via a simulation study. The main objective is to assess whether the conditional and observed-data DICs are able to pick the correct model from which the data are generated. More specifically, simulated data are generated from three models: a constant variance model where observations are drawn independently from $\mathcal{N}(0, \sigma^2)$, the SV model and the SVJ model. Three hundred datasets each comprised of $T = 500$ observations are produced from each of these three models. For each dataset, we estimate the three models by constructing Markov chains of length 20000 after a burn-in period of 1000. To compute the observed-data likelihoods for the two stochastic volatility models, we sample $R = 50$ draws from the importance density at every iteration of the MCMC run.

The parameter values are chosen to be comparable to those estimated from financial daily returns data (measured in decimals). They are also similar to those used in other simulation studies, such as those in Chib, Nardari, and Shephard (2002) and Berg, Meyer, and Yu (2004). In particular, we set $\mu = 0$ for all models. Parameters for the log-volatility transition are set to be $\mu_h = -10$, $\phi_h = 0.97$ and $\omega_h^2 = 0.2^2$ for both the SV and SVJ models. Moreover, parameters for the jump component are selected to be $\kappa = 0.03$ and $\delta = 0.03$. Finally, σ^2 is set so that it is comparable to the variance in the stochastic volatility models: $\sigma^2 = \mathrm{e}^{-\mu} = \mathrm{e}^{-10}$.

The priors discussed in Section 3.1 are considered. We choose the same hyperparameters for parameters that are common across models. Moreover, the hyperparameters are selected so that the implied prior means are similar to the estimates from typical financial daily returns data. In particular, we have $\mu_0 = 0$, $\mu_{h0} = -10$, $V_\mu = V_{\mu_h} = 10$, $\phi_{h0} = 0.97$, $V_{\phi_h} = 0.1^2$, $\nu_h = 5$, $S_h = 0.16$, $k_a = 2$, $k_b = 100$, $\delta_0 = -3.07$ and $V_\delta = 0.149$. These values imply $\mathbb{E}\mu = 0$, $\mathbb{E}\mu_h = -10$, $\mathbb{E}\phi_h = 0.908$, $\mathbb{E}\omega_h^2 = 0.2^2$, $\mathbb{E}\kappa = 0.0196$ and $\mathbb{E}\delta = 0.05$.

In the first experiment, 300 datasets are generated from the constant variance (Const-Var) model. Given each dataset both the conditional and observed-data DICs of the two stochastic volatility models are computed. They are then compared to the (observed-data) DIC of the Const-Var model. Specifically, we subtract the latter DIC from the DICs of both the SV and SVJ models, and the results are reported in Figure 1. Recall that a model is preferred if it has a smaller DIC value. Hence, according to the conditional DIC both the SV and SVJ models are favored relative to the correct model for all the generated datasets. In contrast, for the majority of datasets (94.3% and 100% for the SV and SVJ models, respectively), the observed-data DIC favors the correct model. It is worth noting that among the two stochastic volatility models, the conditional DIC tends to prefer the more complex SVJ model.

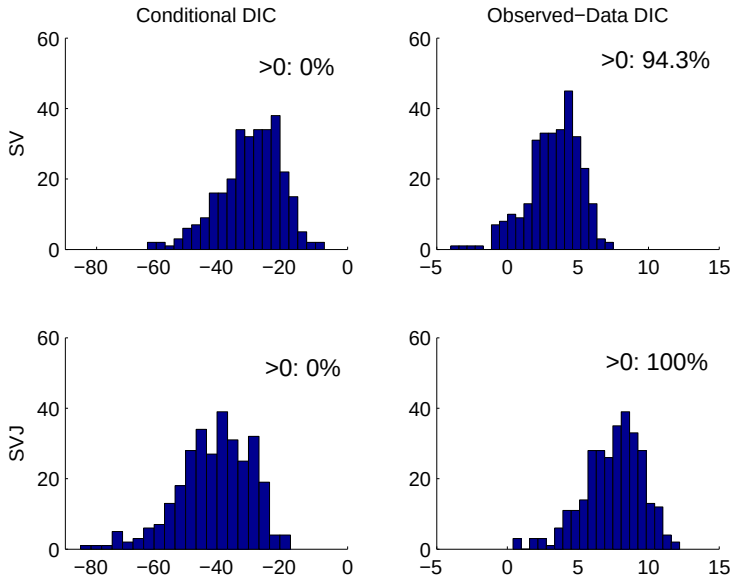

Figure 1: DICs of a given model minus that of the correct model (the Const-Var model). A positive value indicates that the correct model is favored.

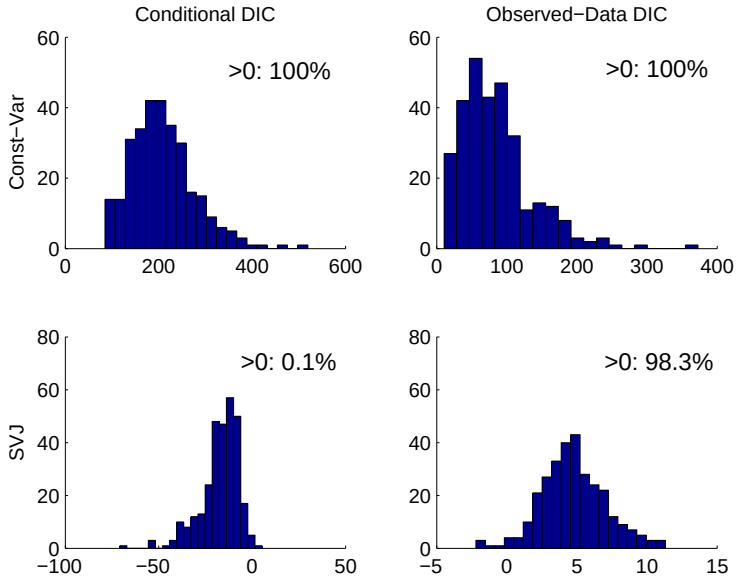

Figure 2: DICs of a given model minus that of the correct model (the SV model). A positive value indicates that the correct model is favored.

In the second experiment, datasets are generated from the SV model, which includes the Const-Var model as a special case and is also nested within the SVJ model. The DICs

relative to the SV model are reported in Figure 2. Both the conditional and observed-data DICs favor the SV model relative to the Const-Var model. However, the conditional DIC tends to favor the overfitted SVJ model: for 99.9% of the datasets the SVJ model is preferred relative to the correct model. In contrast, the observed-data DIC favors the correct model for 98.3% of the datasets. It is also interesting to note that the differences in observed-data DICs between the SV and the SVJ models are small compared to the differences between the SV and Const-Var models, reflecting a small penalty for overfit relative to "underfit"—a model's inability to fit the data well.

In the last simulation experiment, we generate datasets from the SVJ model, which includes both the Const-Var and SV models as special cases. As before, we report the DICs relative to the correct model, and the results are presented in Figure 3. In this example, both the conditional and observed-data DICs tend to pick the correct, more general SVJ model. In particular, comparing between SV and SVJ, the conditional DIC prefers the correct model for 99.3% of the datasets while the figure for the observed-data DIC is 89% of the datasets. It seems that it is relatively more difficult to discriminate between the SV and SVJ models as the jumps occur only rarely (about 3% of the sample).

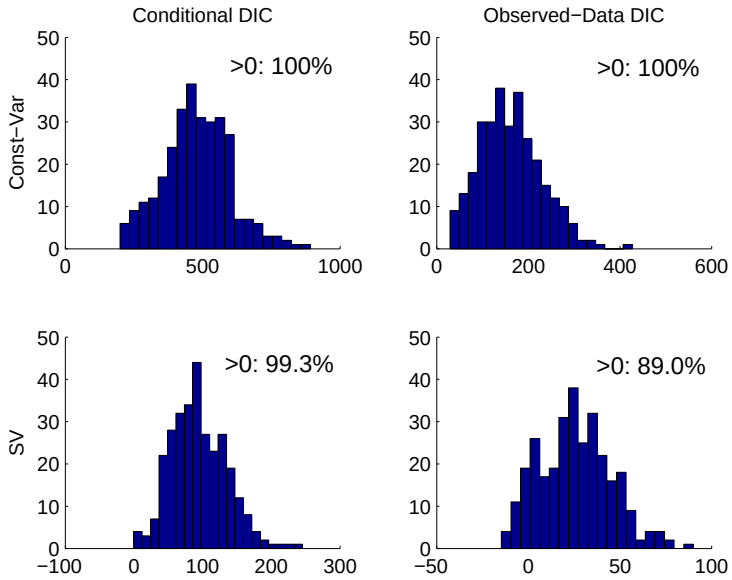

Figure 3: DICs of a given model minus that of the correct model (the SVJ model). A positive value indicates that the correct model is favored.

Overall, this Monte Carlo Study provides evidence that the conditional DIC tends to pick overfitted models whereas the observed-data DIC seems to perform well.

6 An Empirical Application

In this section we illustrate the methodology for estimating the observed-data likelihood with an application that involves the daily returns (in decimals) on the S&P 500 index. The sample period is January 2007 to December 2012, with a total of $T = 1509$ observations. The time series plot of the data is presented in Figure 4. We estimate the stochastic volatility models listed in Table 1 using the S&P 500 data, and we assess which model fits the data best while taking model complexity into account.

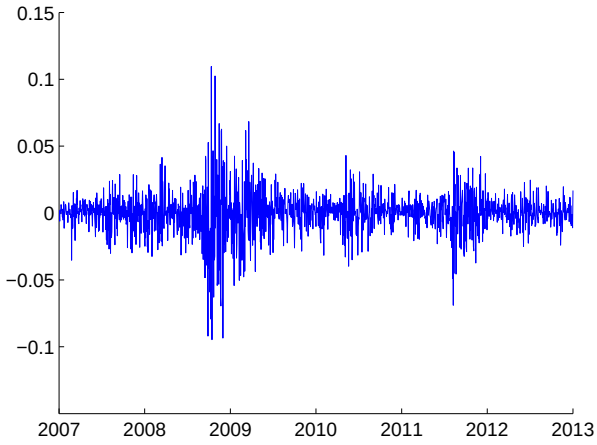

Figure 4: Daily returns on S&P 500 (in decimals) from January 2007 to December 2012.

We use the priors given in Section 3.1 and set the same hyperparameters for parameters that are common across models. For the SV and SVJ models, the same hyperparameters as in the Monte Carlo study in Section 5 are used. For the remaining models, we choose $\rho_{h0} = 0$, $V_{\rho_h} = 1$, $\psi_0 = 0$, $V_\psi = 1$, $\alpha_0 = 0$ and $V_\alpha = 100^2$.

For each model, posterior results are based on 10 parallel chains each of which is of length 10000 after a burn-in period of 1000, with a total of 100000 posterior draws. To compute the observed-data likelihood, we obtain $R = 50$ draws from the importance density at each MCMC iteration. The estimated DICs, their numerical standard errors and computation times (in minutes) are reported in Table 2. The algorithms are implemented using MATLAB on a desktop with an Intel Core i7-870 @2.93 GHz processor.

A few broad conclusions may be drawn. Firstly, the observed-data and conditional DICs favor very different models. For example, the SVMA model is ranked as the best by the observed-data DIC, whereas the conditional DIC ranks it as the worst. In fact, the latter DIC prefers the SVJ model, which is ranked only as the fourth by the former. Hence, erroneous conclusions might be drawn if the conditional DIC favors overfitted models, as suggested by the Monte Carlo study in Section 5.

Table 2: Estimated DICs, numerical standard errors and computation times (in minutes).

	Observed-data DIC	Rank	Time	Conditional DIC	Rank	Time
SV	-9080.8	3	32.7	-9305.0	4	9.4
	(0.56)			(6.18)		
SV2	-9057.5	5	38.1	-9315.4	2	13.6
	(1.20)			(5.55)		
SVJ	-9079.6	4	192.7	-9342.2	1	25.9
	(0.91)			(26.6)		
SVM	-9086.9	2	30.0	-9310.1	3	9.5
	(0.58)			(5.16)		
SVMA	-9087.8	1	55.5	-9296.7	5	20.2
	(0.51)			(5.06)		

Secondly, according to the observed-data DIC, both the volatility feedback (SVM) and moving average (SVMA) components seem to be useful additions to the basic SV model. In contrast, the jump component and the AR(2) transition for the log-volatility are apparently not as useful in modeling the returns on the S&P 500.

Thirdly, the numerical standard errors of the conditional DICs are typically quite large, even after averaging 100000 posterior draws. This highlights the need to report numerical standard errors of the conditional DICs, which is often not done in empirical research. On the other hand, the observed-data DICs are more accurately estimated at a slightly higher computational cost. For instance, when computing the DICs for the SV model, the estimate of the observed-data DIC is about 423 times $((6.18)^2/9.4 \times 32.7/(0.56)^2)$ more accurate—in terms of variance reduction—than that of the conditional DIC after accounting for the computation times.

Next, we report the posterior means and standard deviations of the parameters in Table 3. The parameters governing the transition of the log-volatility have similar estimates across models. In particular, all show high persistence with the posterior mean of ϕ_h estimated to be between 0.95 to 0.986. In addition, an AR(1) transition seems to be sufficient given that the posterior mean of the AR(2) coefficient ρ_h is very small (0.022), which also supports the ranking of the observed-data DIC—it ranks the SV2 model below the SV model (the conditional DIC ranks the SV2 model higher, but the numerical standard errors are too large to be conclusive).

Interestingly, the posterior estimates of κ, α and ψ all seem to support the ranking of the observed-data DIC (but not that of the conditional DIC). For example, recall that when $\psi = 0$, the SVMA model reduces to the SV model. Since the observed-data DIC favors the SVMA model relative to the SV model, one would expect that the posterior distribution of ψ has little mass around zero. In fact, the 95% credible interval of ψ is estimated to be $(-0.126, -0.020)$, which excludes 0. Similarly, when $\alpha = 0$, the SVM model reduces to the SV model. The 95% credible interval of α is estimated to be $(-9.411, -1.158)$, which is consistent with the ranking of the observed-data DIC that favors the SVM model over

the SV model. However, the observed-data DIC does not seem to be able to discriminate between the SV and SVJ models, which is reflected in the small posterior mean of κ relative to its posterior standard deviation.

Table 3: Parameter posterior means and standard deviations (in parenthesis).

	SV	SV2	SVJ	SVM	SVMA
μ	0.0008	0.0009	0.0008	0.0013	0.0008
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)
μ_h	-9.109	-9.161	-9.168	-8.832	-9.113
	(0.431)	(0.567)	(0.477)	(0.967)	(0.438)
ϕ_h	0.985	0.950	0.986	0.984	0.985
	(0.006)	(0.091)	(0.006)	(0.006)	(0.006)
σ_h^2	0.039	0.059	0.037	0.040	0.038
	(0.008)	(0.015)	(0.008)	(0.008)	(0.008)
ρ_h	–	0.022	–	–	–
	–	(0.096)	–	–	–
κ	–	–	0.017	–	–
	–	–	(0.015)	–	–
δ	–	–	0.026	–	–
	–	–	(0.010)	–	–
α	–	–	–	-5.224	–
	–	–	–	(2.097)	–
ψ	–	–	–	–	-0.073
	–	–	–	–	(0.027)

7 Concluding Remarks and Future Research

We have proposed novel importance sampling algorithms for estimating the observed-data likelihoods under a variety of stochastic volatility models, with the goal of computing the observed-data DICs. It is illustrated via a Monte Carlo study that the observed-data DICs based on the proposed importance sampling estimators are able to select the correct model, whereas the conditional DIC tends to favor overfitted models. In the empirical application involving daily returns on the S&P 500, we found that according to the observed-data DIC, both the volatility feedback and moving average components seem to be useful additions to the standard SV model. Moreover, the estimation results support the model ranking of the observed-data DIC but not that of the conditional DIC.

The proposed importance sampling estimators for observed-data likelihoods can be used in other settings, such as for computing the marginal likelihood and for developing more efficient MCMC algorithms (e.g., as an input for particle MCMC methods; see Andrieu, Doucet, and Holenstein 2010). We leave these possibilities for future research. In addition, we have only considered a few popular univariate stochastic volatility models. It would

be useful to develop similar importance sampling algorithms for other univariate and multivariate stochastic volatility models. We leave these extensions for future research.

Appendix A: Estimation Details

In this appendix we provide the estimation details for fitting the stochastic volatility models discussed in Section 3.1.

Standard Stochastic Volatility Model

Section 3.2 presents an outline of a Markov sampler for estimating the standard stochastic volatility model. Here we fill in the details of Step 1: sampling from the conditional density $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$. Following Chan (2014), we first obtain a Gaussian approximation of $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ and use this approximation as a proposal density in the acceptance-rejection Metropolis-Hastings algorithm (see, e.g., Tierney, 1994), where candidate draws are obtained via the precision sampler in Chan and Jeliazkov (2009) instead of Kalman filter-based algorithms.

To approximate $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ using a Gaussian density, note that

$$p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) \propto p(\mathbf{y} \,|\, \mu, \mathbf{h}) p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2).$$

Hence, we first derive explicit expressions for the densities $p(\mathbf{y} \,|\, \mu, \mathbf{h})$ and $p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2)$. It can be shown that the latter density is Gaussian (see, e.g. Chan, 2014). Let $\mathbf{H}_{\phi_h}$ be the following lower triangular matrix:

$$\mathbf{H}_{\phi_h} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_h & 1 & 0 & \cdots & 0 \\ 0 & -\phi_h & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\phi_h & 1 \end{pmatrix}.$$

Then, we have $(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2) \sim \mathcal{N}(\boldsymbol{\delta}_h, (\mathbf{H}_{\phi_h}' \boldsymbol{\Sigma}_h^{-1} \mathbf{H}_{\phi_h})^{-1})$, where $\boldsymbol{\Sigma}_h = \text{diag}(\omega_h^2/(1-\phi_h^2), \omega_h^2, \ldots, \omega_h^2)$ and $\boldsymbol{\delta}_h = \mathbf{H}_{\phi_h}^{-1} \widetilde{\boldsymbol{\delta}}_h$ with $\widetilde{\boldsymbol{\delta}}_h = (\mu_h, (1-\phi_h)\mu_h, \ldots, (1-\phi_h)\mu_h)'$. Hence, its log-density is given by

$$\log p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2) = -\frac{1}{2}(\mathbf{h}' \mathbf{H}_{\phi_h}' \boldsymbol{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \mathbf{h} - 2\mathbf{h}' \mathbf{H}_{\phi_h}' \boldsymbol{\Sigma}_h^{-1} \mathbf{H}_{\phi_h} \boldsymbol{\delta}_h) + c_1, \tag{12}$$

where c_1 is a constant independent of $\mathbf{h}$.

Next, we approximate $p(\mathbf{y} \,|\, \mu, \mathbf{h})$ by a Gaussian density in $\mathbf{h}$. To that end, expand $\log p(\mathbf{y} \,|\, \mu, \mathbf{h}) = \sum_{t=1}^T \log p(y_t \,|\, \mu, h_t)$ around a given point $\widetilde{\mathbf{h}} = (\widetilde{h}_1, \ldots, \widetilde{h}_T)' \in \mathbb{R}^T$ by

a second-order Taylor expansion (the choice of $\widetilde{\mathbf{h}}$ is discussed below):

$$\begin{aligned}\log p(\mathbf{y} \,|\, \mu, \mathbf{h}) &\approx \log p(\mathbf{y} \,|\, \mu, \widetilde{\mathbf{h}}) + (\mathbf{h} - \widetilde{\mathbf{h}})'\mathbf{f} - \frac{1}{2}(\mathbf{h} - \widetilde{\mathbf{h}})'\mathbf{G}(\mathbf{h} - \widetilde{\mathbf{h}}) \\ &= -\frac{1}{2}(\mathbf{h}'\mathbf{G}\mathbf{h} - 2\mathbf{h}'(\mathbf{f} + \mathbf{G}\widetilde{\mathbf{h}})) + c_2, \end{aligned} \tag{13}$$

where c_2 is a constant independent of $\mathbf{h}$, $\mathbf{f} = (f_1, \ldots, f_T)'$ and $\mathbf{G} = \text{diag}(G_1, \ldots, G_T)$ with

$$f_t = \frac{\partial}{\partial h_t} \log p(y_t \,|\, \mu, h_t)|_{h_t = \widetilde{h}_t}, \qquad G_t = -\frac{\partial^2}{\partial h_t^2} \log p(y_t \,|\, \mu, h_t)|_{h_t = \widetilde{h}_t}.$$

That is, $\mathbf{G}$ is the negative Hessian of the log-density evaluated at $\widetilde{\mathbf{h}}$. For the standard stochastic volatility model, $\mathbf{G}$ is diagonal (hence a band matrix). In particular, since the log-density of y_t given μ and h_t is given by

$$\log p(y_t \,|\, \mu, h_t) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{1}{2}\text{e}^{-h_t}(y_t - \mu)^2,$$

it is easy to check that

$$\begin{aligned}\frac{\partial}{\partial h_t} \log p(y_t \,|\, \mu, h_t) &= -\frac{1}{2} + \frac{1}{2}\text{e}^{-h_t}(y_t - \mu)^2, \\ \frac{\partial^2}{\partial h_t^2} \log p(y_t \,|\, \mu, h_t) &= -\frac{1}{2}\text{e}^{-h_t}(y_t - \mu)^2. \end{aligned}$$

Now, combining (12) and (13), we have

$$\begin{aligned}\log p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2) &= \log p(\mathbf{y} \,|\, \mu, \mathbf{h}) + \log p(\mathbf{h} \,|\, \mu_h, \phi_h, \omega_h^2) + c_3, \\ &\approx -\frac{1}{2}(\mathbf{h}'\mathbf{K}_\mathbf{h}\mathbf{h} - 2\mathbf{h}'\mathbf{k}_\mathbf{h}) + c_4, \end{aligned} \tag{14}$$

where c_3 and c_4 are constants independent of $\mathbf{h}$, $\mathbf{K}_\mathbf{h} = \mathbf{H}_{\phi_h}'\boldsymbol{\Sigma}_h^{-1}\mathbf{H}_{\phi_h} + \mathbf{G}$ and $\mathbf{k}_\mathbf{h} = \mathbf{f} + \mathbf{G}\widetilde{\mathbf{h}} + \mathbf{H}_{\phi_h}'\boldsymbol{\Sigma}_h^{-1}\mathbf{H}_{\phi_h}\boldsymbol{\delta}_\mathbf{h}$. The expression in (14) is in fact the log-kernel of the $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_\mathbf{h}^{-1})$ density, where $\widehat{\mathbf{h}} = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k}_\mathbf{h}$ (see, e.g., Kroese and Chan, 2014, p. 238). Therefore, $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ can be approximated by the Gaussian density with mean vector $\widehat{\mathbf{h}}$ and precision matrix $\mathbf{K}_\mathbf{h}$. It is important to note that $\mathbf{K}_\mathbf{h}$ is a band matrix; in fact, its nonzero elements appear only on the main diagonal and the diagonals above and below the main diagonal. Consequently, $\widehat{\mathbf{h}}$ can be computed quickly by solving the linear system $\mathbf{K}_\mathbf{h}\mathbf{x} = \mathbf{k}_\mathbf{h}$ for $\mathbf{x}$, and draws from $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_\mathbf{h}^{-1})$ can be efficiently obtained using the precision sampler in Chan and Jeliazkov (2009). This Gaussian approximation is then used as the proposal density in the acceptance-rejection Metropolis-Hastings algorithm.

Finally, the point $\widetilde{\mathbf{h}}$ used in the Taylor expansion in (13) is chosen to be the mode of $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$, which can be quickly obtained by the Newton-Raphson method (see, e.g., Kroese, Taimre, and Botev, 2011, pp. 688-689). First, note that from (14) it follows that the negative Hessian of $\log p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \omega_h^2)$ evaluated at $\mathbf{h} = \widetilde{\mathbf{h}}$ is $\mathbf{K}_\mathbf{h}$ and the

gradient at $\mathbf{h} = \widetilde{\mathbf{h}}$ is $-\mathbf{K_h}\widetilde{\mathbf{h}} + \mathbf{k_h}$. Hence, we can implement the Newton-Raphson method as follows: initialize with $\mathbf{h} = \mathbf{h}^{(1)}$ for some constant vector $\mathbf{h}^{(1)}$. For $l = 1, 2, \ldots$, use $\widetilde{\mathbf{h}} = \mathbf{h}^{(l)}$ in the evaluation of $\mathbf{K_h}$ and $\mathbf{k_h}$, and compute

$$\mathbf{h}^{(l+1)} = \mathbf{h}^{(l)} + \mathbf{K}_\mathbf{h}^{-1}(-\mathbf{K_h}\mathbf{h}^{(l)} + \mathbf{k_h}) = \mathbf{K}_\mathbf{h}^{-1}\mathbf{k_h}.$$

Repeat this procedure until some convergence criterion is reached, e.g., when $\|\mathbf{h}^{(l+1)} - \mathbf{h}^{(l)}\| < c$ for some prefixed tolerance level c.

Stochastic Volatility Model with AR(2) State Transition

Estimation of this variant with an AR(2) transition equation requires only minor modifications of the main algorithm for the standard stochastic volatility model. Specifically, let $\mathbf{H}_{\boldsymbol{\theta}_h}$ be the following lower triangular matrix:

$$\mathbf{H}_{\boldsymbol{\theta}_h} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ -\rho_h & -\phi_h & 1 & 0 & \cdots & 0 \\ 0 & -\rho_h & -\phi_h & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho_h & -\phi_h & 1 \end{pmatrix}.$$

Then, we can rewrite the state equation of h_t in (5) as:

$$\mathbf{H}_{\boldsymbol{\theta}_h}\mathbf{h} = \widetilde{\boldsymbol{\gamma}}_\mathbf{h} + \boldsymbol{\varepsilon}^h, \qquad \boldsymbol{\varepsilon}^h \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_h)$$

where $\boldsymbol{\varepsilon}^h = (\varepsilon_1^h, \ldots, \varepsilon_T^h)'$, $\widetilde{\boldsymbol{\gamma}}_h = (\mu_h, \mu_h, (1-\phi_h-\rho_h)\mu_h, \ldots, (1-\phi_h-\rho_h)\mu_h)'$ and $\mathbf{P}_h$ is a diagonal matrix in which the first two diagonal elements are the unconditional variance $(1-\rho_h)\omega_h^2(1+\rho_h)^{-1}((1-\rho_h)^2 - \phi_h^2)^{-1}$ and the remaining $T-2$ elements equal ω_h^2. It follows that $(\mathbf{h} \,|\, \mu_h, \phi_h, \rho_h, \omega_h^2) \sim \mathcal{N}(\boldsymbol{\gamma}_h, (\mathbf{H}'_{\boldsymbol{\theta}_h}\mathbf{P}_h^{-1}\mathbf{H}_{\boldsymbol{\theta}_h})^{-1})$, where $\boldsymbol{\gamma}_h = \mathbf{H}_{\boldsymbol{\theta}_h}^{-1}\widetilde{\boldsymbol{\gamma}}_h$. Hence, we have

$$\log p(\mathbf{h} \,|\, \mu_h, \phi_h, \rho_h, \omega_h^2) = -\frac{1}{2}(\mathbf{h}'\mathbf{H}'_{\boldsymbol{\theta}_h}\mathbf{P}_h^{-1}\mathbf{H}_{\boldsymbol{\theta}_h}\mathbf{h} - 2\mathbf{h}'\mathbf{H}'_{\boldsymbol{\theta}_h}\mathbf{P}_h^{-1}\mathbf{H}_{\boldsymbol{\theta}_h}\boldsymbol{\gamma}_h) + c_5, \qquad (15)$$

where c_5 is a constant independent of $\mathbf{h}$. Therefore, we only need to replace (12) by (15), and the main algorithm for the standard stochastic volatility model can be directly applied. Minor modifications to the main algorithm are also needed to sample $\boldsymbol{\theta}_h$, μ_h and σ_h^2.

Stochastic Volatility Model with Jumps

To estimate the stochastic volatility model with jumps, a few modifications of the main algorithm are needed. Firstly, it is easy to see that the first and second derivatives of the

conditional likelihood with respect to h_t are respectively

$$\frac{\partial}{\partial h_t}\log p(y_t \,|\, \mu, k_t, q_t, h_t) = -\frac{1}{2} + \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mu - k_t q_t)^2,$$
$$\frac{\partial^2}{\partial h_t^2}\log p(y_t \,|\, \mu, k_t, q_t, h_t) = -\frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mu - k_t q_t)^2.$$

Then, $\mathbf{h}$ can be sampled as before. Secondly, we need additional steps to sample $\mathbf{k} = (k_1, \ldots, k_T)'$, $\mathbf{q} = (q_1, \ldots, q_T)'$, κ and δ from the appropriate conditional distributions. Following Chib, Nardari, and Shephard (2006), we sample $\mathbf{k}$ and δ jointly as follows. First, let $\zeta_t = \log(1+k_t)$ and stack $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_T)'$ over t. If k_t is small, as is the case for high frequency financial returns that are measured in decimals, $\exp(\zeta_t) \approx 1 + \zeta_t$, which implies $k_t q_t \approx \zeta_t q_t$. Recall that the prior for the jump size k_t is given by $\zeta_t = \log(1+k_t) \sim \mathcal{N}(-0.5\delta^2, \delta^2)$. Hence, we can integrate out ζ_t analytically. This allows us to sample δ marginally of $\boldsymbol{\zeta}$ from the density

$$p(\delta)\prod_{t=1}^{T}\phi(y_t \,|\, \mu - 0.5\delta^2 q_t, \delta^2 q_t^2 + \mathrm{e}^{h_t})$$

by the MH algorithm, where $p(\delta)$ is the prior density of δ and $\phi(x \,|\, a, b)$ is the Gaussian density with mean a and variance b evaluated at x. Once δ is sampled, we can draw $\zeta_1, \ldots, \zeta_T$ sequentially as follows: if q_t is zero, we sample ζ_t from the prior $\mathcal{N}(-0.5\delta^2, \delta^2)$; otherwise we sample from $\mathcal{N}(\widehat{\zeta}_t, D_{\zeta_t})$ where $D_{\zeta_t}^{-1} = \delta^{-2} + \mathrm{e}^{-h_t}$ and $\widehat{\zeta}_t = D_{\zeta_t}(-0.5 + \mathrm{e}^{-h_t}(y_t - \mu))$. Next, note that $q_1, \ldots, q_T$ are conditionally independent given the data and other parameters and they can be sampled individually. In particular, each q_t follows a Bernoulli distribution with

$$\mathbb{P}(q_t = 1 \,|\, y_t, \zeta_t, h_t, \kappa) \propto \kappa\phi(y_t \,|\, \mu + \mathrm{e}^{\zeta_t} - 1, \mathrm{e}^{h_t})$$
$$\mathbb{P}(q_t = 0 \,|\, y_t, \zeta_t, h_t, \kappa) \propto (1-\kappa)\phi(y_t \,|\, \mu, \mathrm{e}^{h_t}).$$

Finally, we sample $(\kappa \,|\, \mathbf{q}) \sim \mathcal{B}(k_a + \sum_{t=1}^{T} q_t, k_b + T - \sum_{t=1}^{T} q_t)$.

Stochastic Volatility in Mean Model

To estimate the stochastic volatility in mean model, we only need to make two modifications of the main algorithm. Firstly, the first and second derivatives of the conditional likelihood with respect to h_t become

$$\frac{\partial}{\partial h_t}\log p(y_t \,|\, \mu, \alpha, h_t) = -\frac{1}{2} - \frac{1}{2}\alpha^2\mathrm{e}^{h_t} + \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mu)^2,$$
$$\frac{\partial^2}{\partial h_t^2}\log p(y_t \,|\, \mu, \alpha, h_t) = -\frac{1}{2}\alpha^2\mathrm{e}^{h_t} - \frac{1}{2}\mathrm{e}^{-h_t}(y_t - \mu)^2.$$

Then, $\mathbf{h}$ can be sampled as before. Secondly, we replace Step 2 of the main algorithm with the joint sampling of (μ, α) from $p(\mu, \alpha \,|\, \mathbf{y}, \mathbf{h}, \mu_h, \phi_h, \omega_h^2) = p(\mu, \alpha \,|\, \mathbf{y}, \mathbf{h})$. To that

end, let $\boldsymbol{\beta} = (\mu, \alpha)'$, $\mathbf{V}_{\boldsymbol{\beta}} = \text{diag}(V_\mu, V_\alpha)$, $\boldsymbol{\beta}_0 = (\mu_0, \alpha_0)'$ and

$$\mathbf{X}_{\boldsymbol{\beta}} = \begin{pmatrix} 1 & \mathrm{e}^{h_1} \\ \vdots & \vdots \\ 1 & \mathrm{e}^{h_T} \end{pmatrix}.$$

Then, by standard results, we have $(\mu, \alpha \,|\, \mathbf{y}, \mathbf{h}) \sim \mathcal{N}(\widehat{\boldsymbol{\beta}}, \mathbf{D}_{\boldsymbol{\beta}})$, where $\mathbf{D}_{\boldsymbol{\beta}}^{-1} = \mathbf{V}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}_{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{X}_{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}} = \mathbf{D}_{\boldsymbol{\beta}}(\mathbf{V}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \mathbf{X}_{\boldsymbol{\beta}}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\mathbf{y})$ with $\boldsymbol{\Sigma}_{\mathbf{y}} = \text{diag}(\mathrm{e}^{h_1}, \ldots, \mathrm{e}^{h_T})$.

Stochastic Volatility Model with MA(1) Errors

A few modifications of the main algorithm are needed to fit this variant with MA(1) errors in the observation equation. Firstly, by appropriately transforming the data, we can sample $\mathbf{h}$ as before. Specifically, let

$$\mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \psi & 1 & 0 & \cdots & 0 \\ 0 & \psi & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \psi & 1 \end{pmatrix}.$$

Then, (9) can be written as

$$\boldsymbol{\varepsilon}^{\mathbf{y}} = \mathbf{H}_\psi \mathbf{u}, \qquad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}}),$$

where $\boldsymbol{\Sigma}_{\mathbf{y}} = \text{diag}(\mathrm{e}^{h_1}, \ldots, \mathrm{e}^{h_T})$. Hence, if we transform the data $\mathbf{y}$ via $\widetilde{\mathbf{y}} = \mathbf{H}_\psi^{-1}(\mathbf{y} - \mu\mathbf{1})$, where $\mathbf{1}$ is a $T \times 1$ column of ones, then $(\widetilde{\mathbf{y}} \,|\, \mathbf{h}, \psi, \mu) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}})$. Therefore, by applying Step 1 to the transformed data $\widetilde{\mathbf{y}}$, we can sample $\mathbf{h}$ as before.

Next, to sample μ, observe that it follows from (8) and (9) that $(\mathbf{y} \,|\, \mu, \mathbf{h}, \psi) \sim \mathcal{N}(\mu\mathbf{1}, \boldsymbol{\Omega}_{\mathbf{y}})$, where $\boldsymbol{\Omega}_{\mathbf{y}} = \mathbf{H}_{\boldsymbol{\psi}}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{H}_{\boldsymbol{\psi}}'$. Note that $\boldsymbol{\Omega}_{\mathbf{y}}$ is a band matrix with only a small number of non-zero elements along the main diagonal band. Consequently, computations involving $\boldsymbol{\Omega}_{\mathbf{y}}$ are fast. For computation details see Chan (2013). By standard linear regression results, we have $(\mu \,|\, \mathbf{y}, \mathbf{h}, \psi) \sim \mathcal{N}(\widehat{\mu}, D_\mu)$, where $D_\mu^{-1} = V_\mu^{-1} + \mathbf{1}'\boldsymbol{\Omega}_{\mathbf{y}}^{-1}\mathbf{1}$ and $\widehat{\mu} = (V_\mu^{-1}\mu_0 + \mathbf{1}'\boldsymbol{\Omega}_{\mathbf{y}}^{-1}\mathbf{y})$. Lastly, we sample ψ using an independence chain Metropolis-Hastings step as described in Chan (2013).

Appendix B: Importance Sampling for Observed-Data Likelihoods

In this appendix we provide the details of the importance sampling algorithms. For the SV2, SVM and SVMA models, the only latent variables are the log-volatilties. For each

of these models, we can use the Gaussian approximation of the conditional density of $\mathbf{h}$ given the data and other parameters as the importance density (see Appendix A for details). For example, under the SV2 model, we replace the prior density in (12) by (15) and the Gaussian approximation of $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \mu_h, \phi_h, \rho_h, \omega_h^2)$ can be obtained following the same procedure as in Section 4. Moreover, all the Gaussian approximations can be quickly evaluated at any point as their precision matrices are all band matrices.

For the SVJ model, one needs to integrate out the log-volatilties $\mathbf{h}$, the jumps $\mathbf{q}$ and the jump sizes $\boldsymbol{\zeta}$ through importance sampling. Specifically, we simulate $\mathbf{h}$, $\mathbf{q}$ and $\boldsymbol{\zeta}$ as below. First, given the current posterior mode $\widehat{\mathbf{h}} = (\widehat{h}_1, \ldots, \widehat{h}_T)'$ and other parameters, we generate each q_t from the Bernoulli distribution with

$$\begin{aligned}\mathbb{P}(q_t = 1) &\propto \kappa \phi(y_t \,|\, \mu - 0.5\delta^2, \delta^2 + \mathrm{e}^{\widehat{h}_t}) \\ \mathbb{P}(q_t = 0) &\propto (1-\kappa)\phi(y_t \,|\, \mu, \mathrm{e}^{\widehat{h}_t}).\end{aligned}$$

Then, given the simulated draw $\mathbf{q}^*$, we draw $\zeta_1, \ldots, \zeta_T$ sequentially as follows: if q_t^* is zero, we sample ζ_t from the prior $\mathcal{N}(-0.5\delta^2, \delta^2)$; otherwise we sample from $\mathcal{N}(\widehat{\zeta}_t, D_{\zeta_t})$ where $D_{\zeta_t}^{-1} = \delta^{-2} + \mathrm{e}^{-\widehat{h}_t}$. Lastly, given $\mathbf{q}^*$ and $\boldsymbol{\zeta}^*$, we generate a draw from the Gaussian approximation of $p(\mathbf{h} \,|\, \mathbf{y}, \mu, \boldsymbol{\zeta}^*, \mathbf{q}^*, \mu_h, \phi_h, \omega_h^2)$, obtained as described in Appendix A. In this case, it is also easy to evaluate the importance density, which is a product of Bernoulli and Gaussian densities.

References

C. A. Abanto-Valle, D. Bandyopadhyay, V. H. Lachos, and I. Enriquez. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis*, 54(12):2883–2898, 2010.

C. Andrieu, A. Doucet, and R. Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society Series B*, 72(3):269–342, 2010.

A. Berg, R. Meyer, and J. Yu. Deviance information criterion for comparing stochastic volatility models. *Journal of Business and Economic Statistics*, 22(1):107–120, 2004.

C. Brooks and M. Prokopczuk. The dynamics of commodity prices. *Quantitative Finance*, 13(4):527–542, 2013.

G. Celeux, F. Forbes, C. P. Robert, and D. M. Titterington. Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–674, 2006.

J. C. C. Chan. Moving average stochastic volatility models with application to inflation forecast. *Journal of Econometrics*, 176(2):162–172, 2013.

J. C. C. Chan. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Working Paper*, 2014.

J. C. C. Chan and A. L. Grant. Fast computation of the deviance information criterion for latent variable models. *CAMA Working Paper 9/2014*, 2014.

J. C. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1:101–120, 2009.

J. C. C. Chan, G. Koop, and S. M. Potter. A new model of trend inflation. *Journal of Business and Economic Statistics*, 31(1):94–106, 2013.

S. Chib, F. Nardari, and N. Shephard. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2):281–316, 2002.

S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006.

T. Cogley and T. J. Sargent. Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005.

B. Djegnéné and W. J. McCausland. The HESSIAN method for models with leverage-like effects. *Journal of Financial Econometrics*, 2014. Forthcoming.

J. Durbin and S. J. Koopman. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84:669–684, 1997.

E. Eisenstat and R. W. Strachan. Modelling inflation volatility. *CAMA Working Paper 24/2014*, 2014.

M. J. Jensen and J. M. Maheu. Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157(2):306–316, 2010.

S. Kim, N. Shepherd, and S. Chib. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65(3):361–393, 1998.

S. J. Koopman and E. Hol Uspensky. The stochastic volatility in mean model: Empirical evidence from international stock markets. *Journal of Applied Econometrics*, 17(6): 667–689, 2002.

D. P. Kroese and J. C. C. Chan. *Statistical Modeling and Computation*. Springer, New York, 2014.

D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. John Wiley and Sons, New York, 2011.

Y. Li, T. Zeng, and J. Yu. Robust deviance information criterion for latent variable models. *SMU Economics and Statistics Working Paper Series*, 2012.

W. J. McCausland. The HESSIAN method: Highly efficient simulation smoothing, in a nutshell. *Journal of Econometrics*, 168(2):189–206, 2012.

W. J. McCausland, S. Miller, and D. Pelletier. Simulation smoothing for state-space models: A computational efficiency analysis. *Computational Statistics and Data Analysis*, 55(1):199–212, 2011.

R. B. Millar. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes factors. *Biometrics*, 65(3):962–969, 2009.

H. Mumtaz and P. Surico. Evolving international inflation dynamics: World and country-specific factors. *Journal of the European Economic Association*, 10(4):716–734, 2012.

H. Mumtaz and F. Zanetti. The impact of the volatility of monetary policy shocks. *Journal of Money, Credit and Banking*, 45(4):535–558, 2013.

J. Nakajima and Y. Omori. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics and Data Analysis*, 56(11):3690–3704, 2012.

M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599, 1999.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72(3):821–852, 2005.

H. Rue. Fast sampling of Gaussian Markov random fields with applications. *Journal of the Royal Statistical Society Series B*, 63(2):325–338, 2001.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace. *Journal of the Royal Statistical Society Series B*, 71(2):319–392, 2009.

D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64(4): 583–639, 2002.

L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994.

G. Tsiotas. On generalised asymmetric stochastic volatility models. *Computational Statistics and Data Analysis*, 56(1):151–172, 2012.

M. Vo. Oil and stock market volatility: A multivariate stochastic volatility perspective. *Energy Economics*, 33(5):956–965, 2011.

J. J. J. Wang, S. T. B. Choy, and J. S. K. Chan. Modelling stochastic volatility using generalized t distribution. *Journal of Statistical Computation and Simulation*, 83(2): 340–354, 2013.

J. Yu and R. Meyer. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25(2-3):361–384, 2006.