On efficient Bayesian inference for models with stochastic volatility

D. K. Sakaria and J. E. Griffin*

Abstract

An efficient method for Bayesian inference in stochastic volatility models uses a linear state space representation to define a Gibbs sampler in which the volatilities are jointly updated. This method involves the choice of an offset parameter and we illustrate how its choice can have an important effect on the posterior inference. A Metropolis-Hastings algorithm is developed to robustify this approach to choice of the offset parameter. The method is illustrated on simulated data with known parameters, the daily log returns of the Eurostoxx index and a Bayesian vector autoregressive model with stochastic volatility.

Keywords: Stochastic volatility; Bayesian methods; Markov chain Monte Carlo; Mixture offset representation.

*Corresponding author: Professor Jim Griffin, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K. (Email: J.E.Griffin-28@kent.ac.uk)

1 Introduction

It is known that the volatility of many economic variables vary over time. Initial work on time-varying volatility often considered asset price returns. Over the long term, the volatility of equity returns may appear to be stable but usually there are periods of high volatility and calm market periods when the volatility may be low (Enders, 2004). Several approaches have been developed to model this time-varying volatility. In ARCH and GARCH models (Engle, 1982; Bollerslev, 1986), the volatility is modeled as a function of the lagged values of the asset returns and the volatility. Alternatively, stochastic volatility models assume that the volatility follows a known stochastic process such as an AR process for the logarithm of volatility (see *e.g.* Harvey and Shephard, 1996).

In this paper, we will concentrate on the Bayesian estimation of stochastic volatility models (see *e.g.* Jacquier et al., 1994; Kim et al., 1998; Chib et al., 2002). The asset returns may be expressed as function of past returns or other economic variables and the log volatility is modeled as a separate AR process. A simple stochastic volatility model assumes that y_t, the log return at time t, can be expressed as

$$\begin{aligned} y_t &= e^{h_t/2}\nu_t, \qquad t = 1, \ldots, T \\ h_t &= \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t. \end{aligned} \tag{1}$$

where ν_t and η_t are independent error terms for which $\nu_t \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$ and $\eta_t \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$, and h_t is the log volatility at time t. The model assumes that the log volatility h_t follows an AR(1) process with parameters μ, ϕ, and σ_η.

Bayesian inference is complicated since this is a non-linear state space model. Several Markov chain Monte Carlo (MCMC) methods have been developed to sample this class of models. Jacquier et al. (1994) used one-at-a-time updating of h_t with a carefully chosen proposal distribution (one-at-a-time or single move updating is

often criticised for higly correlated samples). Samplers which update a block of h_t's often lead to better mixing. For example, Jensen and Maheu (2014) propose to update a block of h_t in an asymmetric, nonparametric stochastic volatility model.

The model (1) can be expressed in linear state space form for h_t using transformed data $\log y_t^2 = h_t + \log \nu_t^2$. Kim et al. (1998) (KSC) approximate the distribution of $\log \nu_t^2$ using a normal mixture distribution leading to a Gaussian linear state space form for h_t conditional on the mixture states for each observation. This allows the volatilities $h_1, \ldots, h_T$ to be updated using Forward Filtering Backward Sampling (FFBS) techniques (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). In order to make the approximation robust for small values of y_t, a small offset parameter c is used and $\log(y_t^2 + c)$ is used in place of $\log y_t^2$ as the transformed data. This leads to samples from an approximate posterior distribution for the parameters of the SV model and $h_1, \ldots, h_T$. KSC suggest an importance sampling scheme for estimating posterior quantities using the approximate posterior as the importance sampling distribution. However, as with any importance sampler, results can become biased if the importance sampling distribution (the approximation) is sufficiently different to the actual posterior distribution. This can be the case if c is poorly chosen. The approach has been developed in various directions. Chib et al. (2002) consider models with Student-t distributed innovation, exogeneous variables and jumps in observations. Omori and Watanabe (2008) consider an asymmetric stochastic volatility model and allow correlation between the returns and the volatility which allows the modelling fo the leverage effect. A multivariate normal approximation is used to express the model in a linear state space form with Gaussian errors. Results show a better performance compared to a single move sampler. More recently, Kastner and Frühwirth-Schnatter (2014) developed centring methods.

More recently, the KSC sampler has been applied to more complicated mod-

els. For example, Belmonte et al. (2013) consider dynamic regression models with stochastic volatility

$$y_t = X_t\beta_t + e^{h_t/2}\nu_t, \qquad t = 1, \ldots, T \tag{2}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t.$$

or Clark (2012) builds a vector autoregressive model with stochastic volatility, which will be further considered in this paper. Let y_t be a $(p \times 1)$-dimensional vector of economics variables and x_t be a $(q \times 1)$-dimensional vector of deterministic variables measured at time t. The data modelled as

$$\Pi(L)(y_t - \Psi x_t) = \epsilon_t \tag{3}$$

where Ψ is a $(p \times q)$-dimensional vector of coefficients, $\Pi(L) = I_p - \Psi_1 L - \Psi_2 L^2 \ldots \Psi_k L^k$ is a lag polynomial and ν_t are independent errors. The errors ϵ_t are modelled using a factor stochastic volatility model. Let A be a lower triangular matrix with 1's on the diagonal then

$$\epsilon_t = A^{-1}\Lambda_t^{0.5}\nu_t, \qquad \nu_t \sim \mathrm{N}(0, I_p),$$
$$\Lambda_t = \mathrm{diag}(e^{h_{1,t}}, e^{h_{2,t}}, \ldots e^{h_{p,t}}),$$
$$h_{i,t}) = h_{i,t-1} + \sigma_{\eta,i}\eta_{i,t}, \qquad \eta_{i,t} \overset{iid}{\sim} \mathrm{N}(0,1) \quad \forall i = 1, 2, \ldots, p.$$

Bayesian inference is made using a Gibbs sampler and the volatilities are updated using the KSC method in the approximate model (*i.e.* using $r_t^\star = \log((y_t - X_t\beta_t)^2 + c)$ for the dynamic regression model or $r_t^\star = A\Psi(L)(y_t - \Psi x_t) + c$ for the vector autoregression model) but other parameters (such as β_t) are updated using the correct (rather than the approximate) stochastic volatity model. Although, this seems to have little effect on inference, the Gibbs sampler is not properly specified. In addition, in these models, the effect of c is harder to understand since the scale of $r_t^\star$ can change substantially between iterations.

This paper makes two main contributions. Firstly, we develop an MCMC framework for sampling from the posterior distribution of the SV model (rather than an approximation to the SV model) using the KSC method as a proposal in a Metropolis-Hastings step for updating the volatilities. Secondly, we introduce a method for specifying the offset parameter using standardisation that robustifies the MCMC algorithm to the scale of the data.

The paper initially considers the problem of sampling the time-varying volatilities in the stochastic volatility model in (1) and considers more complicated models in the examples. The remainder of the paper is organised as follows. Section 2 describes the Kim et al. (1998) method to linearise the log volatility model and the difficulty of using an appropriate value of the offset parameter c is highlighted. In section 3, a standardisation method is introduced and a Metropolis-Hastings (M-H) step is described to propose volatility parameter $h_1, \ldots, h_T$ using Forward Filtering Backward Sampling (FFBS). Results using simulated data, Eurostoxx daily log returns and a vector autoregressive model with stochastic volatility are discussed in section 4. Section 5 concludes.

2 Linear state space method

In this section, we review the sampling method of KSC and illustrate the effect of choosing c. KSC suggest transforming the observations in the SV model in (1) so that

$$\log y_t^2 = h_t + \log \nu_t^2, \tag{4}$$

which is now linear in h_t. The error term $\log \nu_t^2$ has a $\log \chi_1^2$ density, which they show can be accurately approximated by a normal mixture distribution. To make the method robust for small or zero values of y_t, a small offset parameter c is used

(Fuller, 2009) so that y_t^2 is replaced by $y_t^\star = y_t^2 + c$. This leads to the following representation of the model in (4),

$$\log y_t^* = h_t + \log \nu_t^2 \tag{5}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t.$$

KSC suggest approximating the distribution of $z_t = \log \nu_t^2$, which is the $\log \chi_1^2$ distribution, by the density

$$p(z_t) = \sum_{i=1}^{7} q_i f_N(z_t | m_i - 1.2704, v_i^2)$$

where $f_N(x|\mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. Introducing a mixture state indicator s_t for $t = 1, \ldots, T$ allows the model to be written in a linear state space form as

$$\log y_t^* = h_t + m_{s_t} - 1.2704 + v_{s_t} \epsilon_t, \tag{6}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t,$$

where ϵ_t and ν_t are independent for which $\nu_t \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$ and $\eta_t \overset{i.i.d.}{\sim} \mathrm{N}(0,1)$, and $p(s_t = i) = q_i$. This allows a Gibbs sampler to be defined where the parameters of the SV model can be updated from their full conditional distributions and the Gaussian, linear state space form in (6) allowed KSC to block sample the volatility parameter $h = (h_1, h_2, \ldots, h_T)$ using the simulation signal smoother (De Jong and Shephard, 1995). The method samples from an approximate posterior distribution and KSC describe how an importance sampler, with the approximation employed as importance sampling distribution, can be used to estimate posterior quantities of interest.

The choice of c can have a large effect on the inferences. Kim et al. (1998) use $c = 0.001$ for the Mixture Offset parameter, but state that "it is possible to let c depend on the values taken by y_t^2". To illustrate the effect, data were generated

$c = 10^{-n}$		
n	Sample mean	Sample variance
2	5.4019	0.0009
3	3.1361	0.0036
4	1.0933	0.1046
5	-0.2892	0.8586
6	-0.9314	2.3086
7	-1.1706	3.6025
8	-1.2578	4.4397
9	-1.2903	4.8983
10	-1.3022	5.1186

$c = 10^{-n}$		
n	Sample mean	Sample variance
2	-1.0534	2.8484
3	-1.2197	3.9463
4	-1.2728	4.5070
5	-1.2879	4.7261
6	-1.2912	4.7842
7	-1.2917	4.7930
8	-1.2917	4.7940
9	-1.2917	4.7941
10	-1.2917	4.7941

Table 1: Sample mean and variance of $\log(y_t^2 + c) - h_t$ when h is simulated using $\mu = -10, \phi = 0.95, \sigma_\eta^2 = 0.01^2$ (left hand table) and when y_t is simulated directly from $\mathrm{N}(0, 1)$, using different values of the offset parameter c.

using the SV model in (1) with $\mu = -10, \phi = 0.95$ and $\sigma_\eta = 0.01$. The sample mean and sample variance of $\log(y_t^2 + c) - h_t$ for 1500 generated values are shown in Table 1. Since the data is generated, it follows that $\log y_t^2 - h_t$ follows a $\log(\chi_1^2)$ distribution whose mean is -1.2704 and whose variance is 4.93. Difference between these values of the mean and variance of $\log(y_t^2 + c) - h_t$ shows the effect of c. The sample mean and variance are only close to the true values when c is smaller than 10^{-9}. For larger values of c, the sample mean is too large and the sample variance is too small. The value chosen by KSC works well if the log returns are on a unit scale but works badly in this case when $\mu = -10$.

The simulation results suggest that we can only choose a single value of c for all data if we can scale the data appropriately. In more complicated models, for example in a regression setting with stochastic volatility where y_t is modelled as a function of other variables, the residuals could depend on the current values of the parameters in the sampler. In this case, it would be difficult to set one value of c which works with different parameter values. A scaling approach will be developed in Section 3 where we will re-scale y_t to a variable with unit variance. To understand the effect of c on data with a unit variance we simulated 1500 data points from a standard normal distribution and calculated the sample mean and sample variance of $\log(y_t^2 + c)$. The results in Table 1 show that c needs to be near 10^{-3} for the sample mean and sample variance to be near to the values without c.

3 Standardisation approach and its MCMC sampler

The results in the previous section show that the value of the offset parameter c can have a strong influence on the shape of the distribution of $\log(y_t^2 + c)$. Our method is based on the following idea. The model in (1) can be expressed as $\nu_t = y_t e^{-h_t/2}$ where $\nu_t \sim \mathrm{N}(0, 1)$ and so we can choose a single value of c which does not have a strong influence on $\log(\nu_t^2 + c) = \log\left(y_t^2 e^{-h_t} + c\right) = \log(y_t^2 + ce^{h_t}) - h_t$. This suggests the following standardised approximating model

$$\log\left(y_t^2 + ce^{h_t}\right) = h_t + \log \nu_t^2, \tag{7}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_\eta \eta_t.$$

The approach of KSC cannot be directly applied using this representation since it is no longer a linear state space model for h_t. This can be used to define alternative

offset parameterisations by replacing e^{h_t} be an estimator which would allow the standard KSC method to be used. For example, one anonymous referee suggested using

$$y_t^* = \begin{cases} \log\left(y_t^2 + c\,\text{var}(y)\right) & y_t = 0 \\ \log\left(y_t^2 + c\,y_t^2\right) & y_t \neq 0 \end{cases}. \tag{8}$$

However, it is hard to see how this can be easily implemented in more complicated model such as (2) or (3) where KSC is used on a residual rather than observed data. Our approach defines an MCMC sampler for the posterior distribution using the non-linear parameterisation in (1). The approximating model in (7) is used to define a proposal for $h_1, \ldots, h_T$, which can be sampled using FFBS, in a Metropolis-Hastings step. In models such as (2), this scheme allows all model parameters to be correctly sampled using the non-linear parameterisation of the SV model.

To describe the MCMC sampling scheme, we will denote $\theta = \{\mu, \phi, \sigma_\eta\}$ and $h = (h_1, \ldots, h_T)$ and so the joint posterior density for h and θ is given by

$$p(h, \theta|y) = \frac{p(y|h, \theta)p(h|\theta)p(\theta)}{p(y)}.$$

Kim et al. (1998) use an approximate density for $p(y|h, \theta)$ and we denote this by

$$\tilde{p}(y_t|h_t, \theta_t) = \sum_{i=1}^{7} \tilde{p}(s_t = i)\tilde{p}(y_t|h_t, \theta_t, s_t = i) = \sum_{i=1}^{7} q_i \tilde{p}(y_t|h_t, \theta_t, s_t = i)$$

where

$$\tilde{p}(y_t|h_t, \theta_t, s_t) = \text{N}\left(\log(y_t^2 + ce^{h_t})|h_t + m_{s_t} - 1.2704, v_{s_t}^2\right).$$

To allow us to use FFBS to update h, we use an augmented version of the posterior distribution for the non-linear parameterisation

$$p_q(h, \theta, s|y) \propto p(y|h, \theta)p(h|\theta)p(\theta)g(s|h, \theta) \tag{9}$$

where $g(s|h, \theta, y)$ is defined as

$$g(s|h, \theta, y) = \frac{\tilde{p}(y|h, \theta, s)\tilde{p}(s)}{\tilde{p}(y|h, \theta)} \propto \prod_{t=1}^{T} \frac{\tilde{p}(y_t|h_t, \theta, s_t)\tilde{p}(s_t)}{\tilde{p}(y_t|h_t, \theta)} \tag{10}$$

which is the full conditional of s using the Kim et al. (1998) approximation. Since $\sum_s g(s|h_t, \theta, y) = 1$,

$$p_q(h, \theta|y) = \sum_s p_q(h, \theta, s|y) \propto p(y|h, \theta)p(h|\theta)p(\theta) = p(h, \theta|y)$$

and sampling from p_q will lead to draw for θ and h from the posterior using the non-linear parameterisation. An MCMC sampling scheme is run with p_q as the target distribution. The parameters θ are updated from $p_q(\theta|y, h)$ followed by s updated from $p_q(s|\theta, y, h)$ and, finally, h updated from $p_q(h|\theta, y, s)$ using a Metropolis-Hastings step where the proposal arises from FFBS. The full details of the scheme are given below with the following priors: $p(\mu) \propto 1$, $\pi(\phi) \propto \left(\frac{1+\phi}{2}\right)^{a-1}\left(\frac{1-\phi}{2}\right)^{b-1}$, $\sigma_\eta^{-2} \sim \text{Ga}\left(\frac{\sigma_r}{2}, \frac{\text{S}_\sigma}{2}\right)$ where $\text{Ga}(a, b)$ represents a gamma distribution with mean a/b and variance a/b^2.

Sampling μ

The full conditional distribution of μ is $\text{N}(\hat{\mu}, \sigma_\mu^2)$ where

$$\hat{\mu} = \sigma_\mu^2 \left(\frac{(1-\phi^2)}{\sigma_\eta^2} h_1 + \frac{(1-\phi)}{\sigma_\eta^2} \sum_{t=1}^{t=T-1} (h_{t+1} - \phi h_t) \right),$$

and

$$\sigma_\mu^2 = \sigma_\eta^2 \left((T-1)(1-\phi)^2 + (1-\phi^2) \right)^{-1}.$$

Sampling ϕ

The full conditional density for ϕ is proportional to $\pi(\phi)p(y|h, \mu, \phi, \sigma_\eta^2)$. We update this parameter using Metropolis-Hastings random walk whose proposal distribution is a normal truncated to $(-1, 1)$ whose mean is the previous value of ϕ. The variance of the proposal is tuned using an Adaptive Metropolis-Hastings method (Atchade and Rosenthal, 2005) to get an acceptance rate of 23.4%.

Sampling σ_η^2

The full conditional distribution of σ_η^{-2} is Ga(a, b) where

$$a = \frac{T + \sigma_r}{2}$$

and

$$b = \frac{\mathrm{S}_\sigma + (h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^{t=T-1}((h_{t+1} - \mu) - \phi(h_t - \mu))^2}{2}.$$

Sampling s

The full conditional distribution of s is $g(s|h, \theta, y)$ and so $s_1, \ldots, s_T$ are conditionally independent with s_t sampled from the discrete distribution

$$p_q(s_t = i|h, \theta, y) = \frac{q_i\, \tilde{p}(\log\left(y_t^2 + ce^{h_t}\right)|h_t, \theta, s_t = i\right)}{\sum_{j=1}^{j=7} q_j\, \tilde{p}\left(\log(y_t^2 + ce^{h_t})|h_t, \theta, s_t = j\right)}.$$

Sampling h

The full conditional density for h is proportional to $p(y|h, \theta)p(h|\theta)g(s|h, \theta, y)$. The parameter is updated using a Metropolis-Hastings step. Suppose that the previous value is h and the proposed value is h' then h' is generated using FFBS on the Gaussian, linear state space model

$$\log(y_t^2 + ce^{h_t}) = h'_t + m_{s_t} - 1.2704 + v_{s_t}\epsilon_t \tag{11}$$

$$h'_t = \mu + \phi(h'_{t-1} - \mu) + \sigma_\eta \eta_t$$

where $\epsilon_t \sim \mathrm{N}(0, 1)$. This is the approximating model in (7) conditioned on $s_1, \ldots, s_T$ and with $\log(y_t^2 + ce^{h_t})$ evaluated at the previous value of h_t. The form of the FFBS steps are available from Carter and Kohn (1994) or Frühwirth-Schnatter (1994). Let $\mu_{B,t}$ and $\sigma_{B,t}^2$ be the mean and variance of the state in the backward steps and μ_t and σ_t^2 are the *a posteriori* mean and variance at time t derived from the Kalman filter.

Then $\mu_{B,T} = \mu_t$ and $\sigma^2_{B,T} = \sigma^2_T$ and

$$\sigma^2_{B,t}|h'_{t+1:T}, h = \left(\frac{1}{\sigma^2_t} + \frac{\phi^2}{\sigma^2_\eta}\right)^{-1},$$

$$\mu_{B,t}|h'_{t+1:T}, h = \left(\frac{1}{\sigma^2_{t-1}} + \frac{\phi^2}{\sigma^2_\eta}\right)^{-1} \left(\frac{\mu_t}{\sigma^2_t} + \frac{\phi h_{t+1}{}' - \phi\mu(1-\phi)}{\sigma^2_\eta}\right),$$

for $t < T$ where $h_t{}'$ is the sample value at time t. The proposal density is

$$q(h'|h, \theta, y, s) = \prod_{t=1}^{T} \mathrm{N}\left(h'_t|\mu_{B,t}, \sigma^2_{B,t}\right).$$

It is important to note that the proposal depends on h_t through the transformed data $\log(y_t^2 + ce^{h_t})$. The Metropolis-Hastings acceptance probability is

$$\begin{aligned} a &= \min\left(1, \frac{p_q(h', \theta, s|y)}{p_q(h, \theta, s|y)} \frac{q(h|h', \theta, y, s)}{q(h'|h, \theta, y, s)}\right) \\ &= \min\left(1, \frac{p(y|h', \theta)p(h'|\theta)g(s|h', \theta, y)}{p(y|h, \theta)p(h|\theta)g(s|h, \theta, y)} \frac{q(h|h', \theta, y, s)}{q(h'|h, \theta, y, s)}\right) \\ &= \min\left(1, \prod_{t=1}^{T} \frac{p(y_t|h'_t, \theta)\tilde{p}(y_t|h_t, \theta)}{p(y_t|h_t, \theta)\tilde{p}(y_t|h'_t, \theta)}\right). \end{aligned}$$

Since, $\tilde{p}(y_t|h_t, \theta)$ is a good approximation to $p(y_t|h_t, \theta)$, we have that

$$\frac{p(y_t|h'_t, \theta)\tilde{p}(y_t|h_t, \theta)}{p(y_t|h_t, \theta)\tilde{p}(y_t|h'_t, \theta)} \approx 1.$$

This suggest that this sampling step will have good acceptance rate for suitably chosen values of T but that its performance will deteriorate as T increases. The amount by which the performance deteriorate will depend on the data.

4 Results

The performances of our proposed method described in Section 3, which we will call the Metropolis-Hastings sampler with standardisation (MH-S) method, and the

KSC method with importance sampling (KSC) were compared with different values of c on simulated and financial data. The priors for the parameters were chosen as in section 3 with the hyperparameter values: $a = 20$, $b = 1.5$ (which gives a prior mean for ϕ of 0.86), $\sigma_r = 5$, and $S_\sigma = 0.01\sigma_r$. The performances were compared using the Effective Sample Size (ESS) (Sokal, 1997) which was estimated by

$$\text{ESS} = \frac{N}{1 + 2\sum_{i=1}^{j} r_i}$$

where N is the number of samples from the iterations, r_i is the correlation coefficient at lag i and j is the count of the non-zero correlation coefficients. The number of lags to include was chosen using Bartlett's test.

4.1 Simulated data

Two test data sets of length 1500 were generated using the model in (1) with $\mu = -10$ and $\phi = 0.95$. One data set used $\sigma_\eta = 0.2$ and the other used $\sigma_\eta = 0.6$. An initial 20 000 iterations were used as a burn in period followed by a further 20 000 iterations which were thinned to 1 in 5 to sample the parameters. The chains were found to be sufficiently long for the trace plots of parameters to stabilise.

The estimated posterior medians and 95% credible intervals of the model parameters for the two simulated data sets using the MH-S and KSC samplers are presented in Table 2. The MH-S sampler clearly provides estimates which are robust to the choice of c over the range of values considered with both data sets. In contrast, the results using the KSC sampler depend on the value of c with both data sets. The larger values of c ($c = 10^{-3}$ and $c = 10^{-5}$) provide extremely biased estimates for all summaries in both data sets. The method performs well with smaller values of c ($c = 10^{-7}$ and $c = 10^{-9}$). This is not surprising since the mean of y_t^2 is $\exp\{\mu + \sigma_\eta^2/2\}$ which is 4.6317×10^{-5} in this example and so is much smaller than the values considered by KSC (where the mean of y_t^2 has a scale of units). If we adjust for the

$\sigma_\eta = 0.2$

$c = 10^{-n}$	Sampler	μ	ϕ	σ_η^2
$n = 3$	MH-S	-10.17 *(-10.39,-9.99)*	0.928 (0.887, 0.958)	0.0508 *(0.0290, 0.0847)*
$n = 5$	MH-S	-10.17 *(-10.36,-9.99)*	0.927 (0.876,0.959)	0.0522 *(0.0299,0.1019)*
$n = 7$	MH-S	-10.17 *(-10.35, -9.98)*	0.929 (0.890,0.958)	0.0504 *(0.0269,0.0814)*
$n = 9$	MH-S	-10.17 *(-10.35, -9.99)*	0.923 (0.875,0.956)	0.0544 *(0.0302,0.0936)*
$n = 3$	KSC	-7.04 *(-7.09,-6.98)*	0.617 *(0.315, 0.836)*	0.0036 *(0.0020, 0.0073)*
$n = 5$	KSC	-9.88 *(-10.03,-9.74)*	0.946 *(0.905,0.972)*	0.0161 *(0.0084,0.0308)*
$n = 7$	KSC	-10.15 *(-10.33, -9.97)*	0.926 *(0.881,0.959)*	0.0515 *(0.0286,0.0914)*
$n = 9$	KSC	-10.17 *(-10.35, -9.99)*	0.924 *(0.874,0.957)*	0.0554 *(0.0298,0.0987)*

$\sigma_\eta = 0.6$

$c = 10^{-n}$	Sampler	μ	ϕ	σ_η^2
$n = 3$	MH-S	-10.17 *(-11.42,-8.89)*	0.979 *(0.967, 0.990)*	0.2010 *(0.1501, 0.2659)*
$n = 5$	MH-S	-10.19 *(-11.46,-8.96)*	0.980 *(0.967,0.990)*	0.2004 *(0.1495,0.2721)*
$n = 7$	MH-S	-10.19 *(-11.43, -8.86)*	0.980 *(0.968,0.991)*	0.1949 *(0.1509,0.2621)*
$n = 9$	MH-S	-10.18 *(-11.37, -8.89)*	0.979 *(0.967,0.990)*	0.2027 *(0.1560,0.2051)*
$n = 3$	KSC	-6.80 *(-7.04,-6.57)*	0.978 *(0.958, 0.991)*	0.0074 *(0.0043, 0.0120)*
$n = 5$	KSC	-9.55 *(-10.49,-8.60)*	0.983 *(0.972,0.992)*	0.0803 *(0.0606,0.1063)*
$n = 7$	KSC	-10.15 *(-11.36, -9.00)*	0.980 *(0.968,0.990)*	0.1870 *(0.1411,0.2462)*
$n = 9$	KSC	-10.18 *(-11.37, -8.93)*	0.979 *(0.966,0.990)*	0.2069 *(0.1560,0.2763)*

Table 2: Posterior medians and 95% credible intervals of the model parameters with data simulated using different values of c.

difference in the scales, an appropriate value would be $c = 4.6317 \times 10^{-8}$ which is contained in the range of vales for c for which the KSC method performs well.

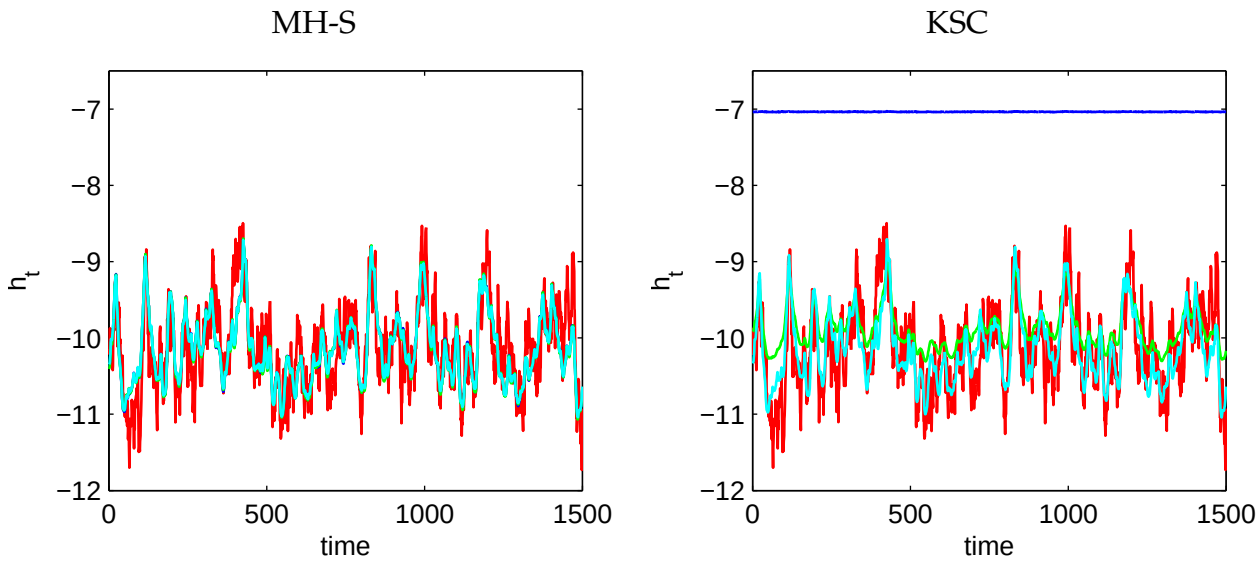

Figure 1: Posterior mean of h_t using the MH-S and KSC methods and different values of c. The colour key for the line plots is: simulated (red), $c = 10^{-3}$ (blue), $c = 10^{-5}$ (green), $c = 10^{-7}$ (light blue).

The KSC method should provide unbiased estimates of all parameters and so the biases in the posterior summaries for $c = 10^{-3}$ and $c = 10^{-5}$ are surprising. The plots in Figure 1 show the posterior means of h_t for different samplers. Clearly, the posterior mean of h_t for the KSC sampler with $c = 10^{-3}$ is larger than the correct posterior values. This is directly due to the choice of c which concentrates the posterior for h_t on larger values and so biases the results for μ. The importance sampler should correct for differences but, in this case, the importance sampling distribution places negligible mass on the correct values and so leads to biased posterior summaries. To confirm that this is the cause, we ran a Metropolis-Hastings chain without standardisation (*i.e.* where (6) is replaced by $\log(y_t^2 + c) = h_t + \log \nu_t^2$)) (MH-

$c = 10^{-n}$	μ	ϕ	σ_η^2
$n = 3$	-7.64 *(-7,72, -7.58)*	0.493 *(0.292, 0.665)*	0.0033 *(0.0025, 0.0058)*
$n = 5$	-6.87 *(-6.85, -2.84)*	0.949 *(0.398,0.967)*	0.0057 *(0.0035,0.061)*
$n = 7$	-10.17 *(-10.35, -9.99)*	0.925 *(0.876,0.957)*	0.0539 *(0.0297,0.0960)*
$n = 9$	-10.17 *(-10.35, -9.98)*	0.929 *(0.888,0.961)*	0.0499 *(0.0274,0.0857)*

Table 3: Posterior medians and 95% credible intervals of the model parameters with data simulated with $\sigma_\eta = 0.2$ and using the MH-NS sampler with different values of c.

NS). Table 3 shows that this sampler also leads to biased estimates for small c. This is an independence Metropolis-Hastings sampler which has a proposal supporting values far from the correct values of h_t when $c = 10^{-3}$ and so the sampler never moves to the correct values of h_t.

The effective sample sizes for the two simulated data set using the MH-S and KSC samplers are presented in Table 4. These show that the MH-S sampler perform well for all values of c despite the length of the time series ($T = 1500$). The sampler has a large number of accepted h moves, good effective sample sizes for h_{100} and μ, and acceptable effective sample sizes for the other model parameters. The effective sample sizes with the KSC sampler are consistently larger for values of c which give unbiased estimates of the posterior summaries ($c = 10^{-7}$ and $c = 10^{-9}$) .

Overall, these results show that the MH-S method is robust to the choice of c and can provide effective inference for relatively long time series ($T = 1500$) with realistic values of the ϕ and σ_η.

$\sigma_\eta = 0.2$

$c = 10^{-n}$	Sampler	h_{100}	μ	ϕ	σ_η^2	h moves (%)
$n = 3$	MH-S	1577.9	2952.9	143.9	76.2	40.5
$n = 5$	MH-S	1068.5	2230.3	96.8	58.9	27.2
$n = 7$	MH-S	1496.9	2625.1	76.6	76.6	41.2
$n = 9$	MH-S	1327.2	2050.1	67.5	67.5	29.0
$n = 3$	KSC	1672.1	181.4	36.0	109.5	
$n = 5$	KSC	3055.6	3371.0	203.8	108.1	
$n = 7$	KSC	3336.9	3430.8	210.0	138.5	
$n = 9$	KSC	2590.7	3323.6	207.9	126.4	

$\sigma_\eta = 0.6$

$c = 10^{-n}$	Sampler	h_{100}	μ	ϕ	σ_η^2	h moves (%)
$n = 3$	MH-S	693.1	3542.4	687.4	172.3	28.0
$n = 5$	MH-S	720.9	4092.8	465.6	126.3	25.5
$n = 7$	MH-S	860.9	3740.2	753.0	217.2	27.4
$n = 9$	MH-S	712.4	4209.0	950.0	238.1	25.3
$n = 3$	KSC	3392.3	3520.5	446.9	202.4	
$n = 5$	KSC	2839.1	4135.3	1438.1	523.5	
$n = 7$	KSC	2715.0	4103.0	1183.8	583.1	
$n = 9$	KSC	3261.3	4016.6	1169.9	529.6	

Table 4: Effective sample size for all model parameters with data simulated using different values of c.

4.2 Eurostoxx Index

The returns of the Eurostoxx index from 2 January 2007 to 23 April 2013 were calculated using the closing values and are shown in Figure 2. There are 1585 log returns

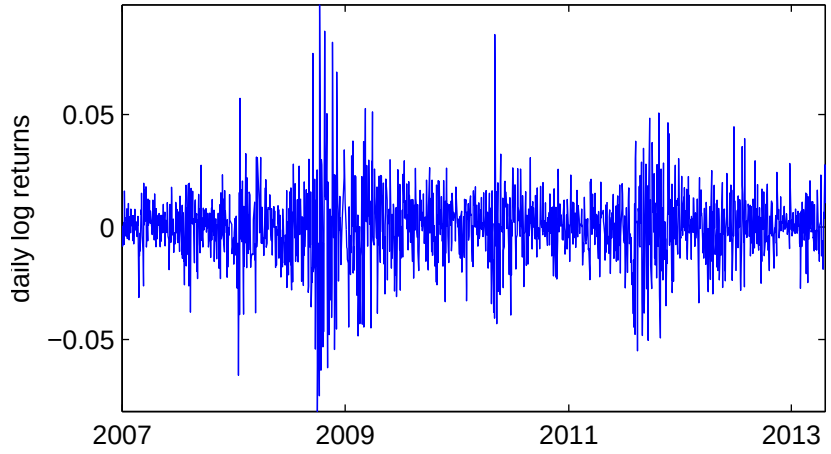

Figure 2: Daily log returns of the Eurostoxx index from January 2007 to April 2013.

in the time series with 3 cases of zero log returns (on 6 January 2010, 3 March 2010 and 16 July 2012). The MH-S and KSC methods were run for different values of c. An initial 20 000 iterations were used as a burn in period followed by a further 20 000 iterations which were thinned to 1 in 5 to sample the parameters.

The estimated posterior median and 95% credible interval for this data set using the MH-S and KSC samplers are presented in Table 5. Again, these results show that the MH-S sampler provides good estimates of the posterior summaries for all value of c whereas the KSC sampler does not provide good summaries for $c = 10^{-3}$. Again, this emphasises that the KSC sampler relies on a value of c which is consistent with the scale of the data whereas the MH-S sampler works well regardless of the scale of the data.

Table 6 shows the ESS for both the MH-S and KSC samplers. The results are consistent with the simulated examples with the KSC sampler having a larger ESS than the MH-S but the MH-S providing a suitable ESS for effective inference.

The KSC sampler was also applied with an alternative offset specification in (8).

$c = 10^{-n}$	Sampler	μ	ϕ	σ_η^2
$n = 3$	MH-S	-8.76 *(-9.32, -8.24)*	0.981 *(0.966, 0.992)*	0.0287 *(0.0178, 0.0504)*
$n = 5$	MH-S	-8.76 *(-9.30, -8.26)*	0.982 *(0.966,0.922)*	0.0284 *(0.0174,0.0496)*
$n = 7$	MH-S	-8.77 *(-9.32, -8.28)*	0.981 *(0.966,0.993)*	0.0286 *(0.0174,0.0480)*
$n = 9$	MH-S	-8.75 *(-9.30, -8.24)*	0.981 *(0.965,0.992)*	0.0282 *(0.0176,0.0493)*
$n = 3$	KSC	-6.84 *(-6.91,-6.77)*	0.768 *(0.417, 0.949)*	0.0045 *(0.0023, 0.0095)*
$n = 5$	KSC	-8.68 *(-9.20,-8.21)*	0.983 *(0.969,0.994)*	0.0218 *(0.0133,0.0355)*
$n = 7$	KSC	-8.76 *(-9.27, -8.28)*	0.980 *(0.964,0.992)*	0.0299 *(0.0172,0.0491)*
$n = 9$	KSC	-8.77 *(-9.28, -8.30)*	0.980 *(0.964,0.991)*	0.0306 *(0.0189,0.0508)*

Table 5: Posterior medians and 95% credible intervals of the model parameters with the Eurostoxx data and using different values of c.

$c = 10^{-n}$	Sampler	h_{100}	μ	ϕ	σ_η^2	h moves (%)
$n = 3$	MH-S	1709.4	3863.5	320.9	118.1	45.2
$n = 5$	MH-S	1772.0	3793.5	202.1	102.4	42.5
$n = 7$	MH-S	1771.6	3398.6	217.7	95.4	43.2
$n = 9$	MH-S	1253.8	3816.7	122.1	56.2	37.0
$n = 3$	KSC	3572.2	409.4	44.0	83.9	
$n = 5$	KSC	3452.8	4001.1	572.1	217.7	
$n = 7$	KSC	2823.0	3805.2	456.9	213.2	
$n = 9$	KSC	3176.6	2960.0	347.0	157.5	

Table 6: Effective sample size for all model parameters with the Eurostoxx data and using different values of c.

The posterior medians of h_t is shown in Figure 3. The results from using a simpler

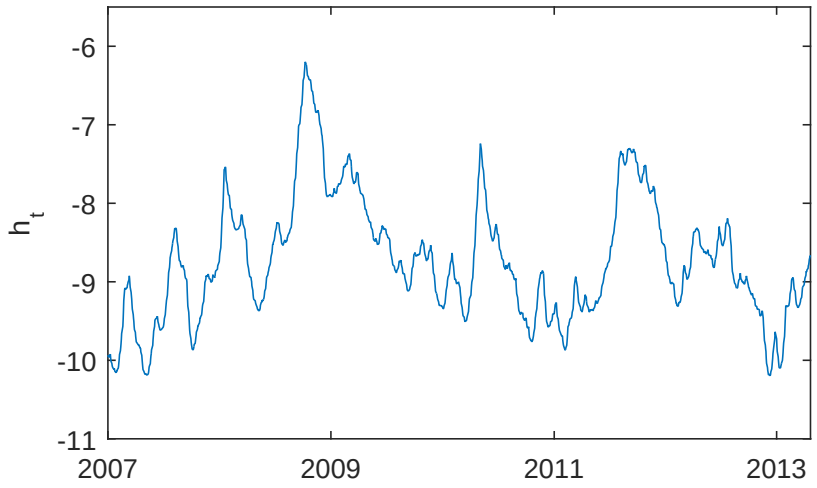

Figure 3: Posterior median of h_t for the Eurostoxx data using the approximate variance to standardise the data.

standandardisation method are similar to the results shown in Figure 4 where the MH-S method was used.

The posterior medians of h_t (with the 95% credible interval) are shown in Figure 4. The volatility shows substantial persisetence and time-variation with higher values during the financial crisis.

4.3 Vector autoregression

The use of stochastic volatility in multivariate time series models has become increasingly common. As an example, we consider the following Bayesian VAR model with stochastic volatility which was introduced in Clark (2012) building on steady-state priors (Villani, 2009). Let y_t be a $(p \times 1)$-dimensional vector of economics variables and x_t be a $(q \times 1)$-dimensional vector of deterministic exogeneous variables

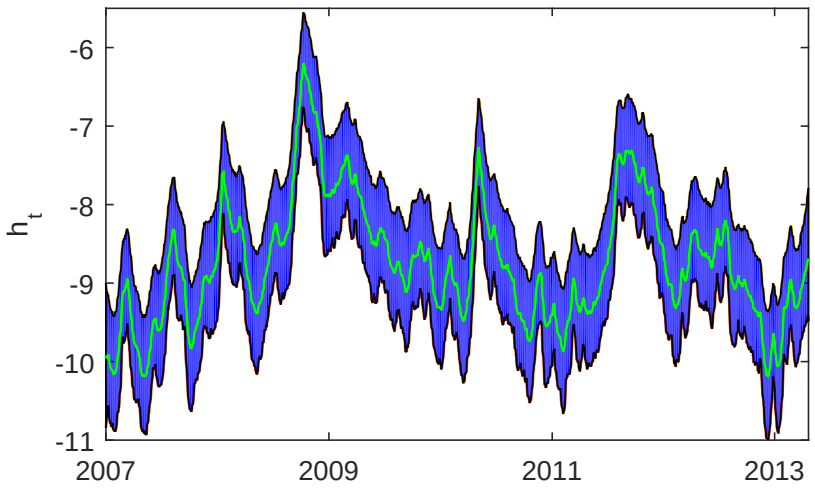

Figure 4: Posterior median and 95% credible interval of h_t for the Eurostoxx data using the MH-S with $c = 10^{-3}$.

measured at time t. The data modelled as

$$\Pi(L)(y_t - \Psi x_t) = \epsilon_t$$

where Ψ is a $(p \times q)$-dimensional vector of coefficients, $\Pi(L) = I_p - \Psi_1 L - \Psi_2 L^2 \dots \Psi_k L^k$ is a lag polynomial and ν_t are independent errors. The vector Ψx_t allows for a deterministic trend within the model. The errors ϵ_t are modelled using a factor stochastic volatility model. Let A be a lower triangular matrix with 1's on the diagonal then

$$\begin{aligned}
\epsilon_t &= A^{-1}\Lambda_t^{0.5}\nu_t, \qquad \nu_t \sim \mathrm{N}(0, I_p), \\
\Lambda_t &= \mathrm{diag}(e^{h_{1,t}}, e^{h_{2,t}}, \dots e^{h_{p,t}}), \\
h_{i,t} &= h_{i,t-1} + \sigma_{\eta,i}\,\eta_{i,t}, \qquad \eta_{i,t} \overset{iid}{\sim} \mathrm{N}(0,1) \quad \forall i = 1, 2, \dots, p.
\end{aligned}$$

Clark (2012) applies the model to output growth, unemployment rate, inflation and federal funds rate with long-term inflation expectations used as an exogeneous

variable and uses real-time data from 1961:Q1 to 2008:Q3. We consider the same data and time period but use the final revision values for all data.

Clark (2012) describes, in detail, the prior specification and the Gibbs sampling steps needed to fit the model and these are not repeated here. The full conditional distribution for each volatility process can be put in the form of a stochastic volatility model by noting that

$$A\Pi(L)(y_t - \Psi x_t) \equiv \tilde{y}_t = \Lambda_t^{0.5}\nu_t$$

and

$$\log \tilde{y}_{i,t}^2 = h_{i,t} + \log \nu_{i,t}^2, \qquad \forall i = 1, \ldots, p.$$

This allows the KSC method described in Section 2 to be used to update $h_{i,1}, \ldots, h_{i,T}$ from their full conditional distribution for each $i = 1, \ldots, p$ in a Gibbs sampler. However, Clark (2012) notes that "in preliminary investigations with BVAR models, estimates based on the latter algorithm seemed to be unduly dependent on the priors and prone to yielding highly variable estimates of the volatility". Therefore, we consider the MH-S algorithm described in Section 3 and show that this algorithm is able to successfully sample from the posterior distribution. The posterior distribution of the model was sampled using 3 different values of the mixture offset ($c = 10^{-3}$, $c = 10^{-6}$ and $c = 10^{-9}$) with the sampler from for a total of 50 000 iterations with a burn-in period of 10 000 iterations. The results were found to be very similar for all parameters and very similar to results from a Metropolis-Hastings algorithm which updates each $h_{i,t}$ separately without taking any transformations. For iilustration, the posterior median of $e^{h_{1,t}}$ for the three values of c are shown in Figure 5 and clearly do not depend on the value of the mixture offset parameter.

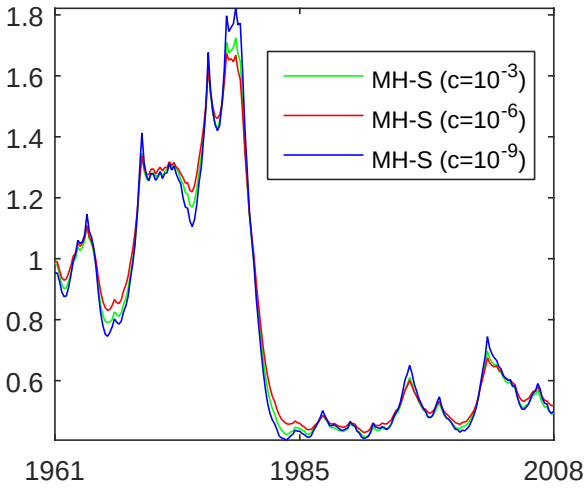

Figure 5: Posterior median of $e^{h_{1,t}}$ for Example 4.3 using different mixture offset values and the MH-S method.

5 Conclusion

When the SV model is expressed in linear state space form using a normal mixture model, the value of the offset parameter c used for the MCMC sampling can have an important effect on the posterior inferences. To overcome this lack of robustness to the choice of c, we propose a Metropolis-Hastings sampler which uses a linear state space constructed using a standardised version of the error term. The volatilities are sampled jointly using a forward filtering backwards sampling algorithms in the same way as KSC. This approach provides inference about the volatility and the model parameters which is robust to the choice of c The effective sample sizes for the volatility and the model parameters indicate that the method can provide accurate inference for realistic values of the model parameters on time series of realistic length.

In contrast to the KSC method, the MCMC chain is generated from the correct full

conditional joint distribution for the volatilities and so no importance sampling reweighting is needed to correct inference. The application to a vector autoregression with stochastic volatility in section 4.3 shows that sampling from the correct full conditional distribution can have important consequences for the ability of the Gibbs sampler to simulate from the correct posterior distribution.

References

Atchade, Y. and S. Rosenthal (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli 11*(5), 815–828.

Belmonte, M. A. G., G. Koop, and D. Korobilis (2013). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting 33*, 80–94.

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

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–553.

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

Clark, T. E. (2012). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*.

De Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*(2), 339–350.

Enders, W. (2004). Applied Econometric Time Series, 2004. *NY: John Wiley & Sons*.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica 50*(4), pp. 987–1007.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183–202.

Fuller, W. A. (2009). *Introduction to statistical time series*, Volume 428. John Wiley & Sons.

Harvey, A. C. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics 14*(4), 429–434.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics 12*, 371–389.

Jensen, M. J. and J. M. Maheu (2014). Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. *Journal of Econometrics 178*, 523–538.

Kastner, G. and S. Frühwirth-Schnatter (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis 76*, 408–423.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies 65*(3), pp. 361–393.

Omori, Y. and T. Watanabe (2008). Block sampler and posterior mode estimation for asymmetric stochastic volatility models. *Computational Statistics & Data Analysis 52*(6), 2892–2910.

Sokal, A. (1997). Monte Carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration*, pp. 131–192. Springer.

Villani, M. (2009). Steady-state priors for vector autoregressions. *Journal of Applied Econometrics 24*, 630–650.

A Derivation of the Metropolis-Hastings acceptance ratio for sampling h

The Metropolis-Hastings acceptance ratio is

$$\begin{aligned} a &= \min\left(1, \frac{p_q(h^{'}, \theta, s|y)}{p_q(h, \theta, s|y)} \frac{q(h|h^{'}, \theta, y, s)}{q(h^{'}|h, \theta, y, s)}\right), \\ &= \min\left(1, \frac{p(y|h^{'}, \theta)p(h^{'}|\theta)g(s|h^{'}, \theta, y)}{p(y|h, \theta)p(h|\theta)g(s|h, \theta, y)} \frac{q(h|h^{'}, \theta, y, s)}{q(h^{'}|h, \theta, y, s)}\right). \end{aligned}$$

Using

$$q(h|h', \theta, y, s) = \frac{\tilde{p}(y|h, \theta, s)p(h|\theta)}{\tilde{p}(y|s, \theta)},$$

and

$$g(s|h', \theta, y) = \frac{\tilde{p}(y|h', \theta, s)\tilde{p}(s)}{\tilde{p}(y|h', \theta)},$$

implies that

$$a = \min\left(1, \frac{p(y|h', \theta)p(h'|\theta)}{p(y|h, \theta)p(h|\theta)} \frac{\frac{\tilde{p}(y|h', \theta, s)\tilde{p}(s)}{\tilde{p}(y|h', \theta)}}{\frac{\tilde{p}(y|h, \theta, s)\tilde{p}(s)}{\tilde{p}(y|h, \theta)}} \frac{\frac{\tilde{p}(h, \theta, s)p(h|\theta)}{\tilde{p}(y|s, \theta)}}{\frac{\tilde{p}(y|h', \theta, s)p(h'|\theta)}{\tilde{p}(y|s, \theta)}}\right).$$

Finally, cancelling out the terms, gives

$$a = \min\left(1, \frac{\frac{p(y|h, \theta)}{\tilde{p}(y|h', \theta)}}{\frac{p(y|h, \theta)}{\tilde{p}(y|h, \theta)}}\right).$$