

# A non-iterative (trivial) method for posterior inference in stochastic volatility models

Mike G. Tsionas\*

January 8, 2017

## Abstract

We propose a new non-iterative, very simple but accurate, Bayesian inference procedure for the stochastic volatility model. The only requirement of our approach is to solve a large, sparse linear system which we avoid by iteration.

**Key words:** Stochastic volatility model; Monte Carlo methods; Markov Chain Monte Carlo; iterative methods.

**Acknowledgments:** The author is indebted to an Associate Editor for many useful comments on an earlier version of the paper.

---

\*Lancaster University Management School, LA1 4YX, UK, m.tsionas@lancaster.ac.uk

# 1 Introduction

Suppose  $\theta \in \Theta \subseteq \mathfrak{R}^k$  is a parameter, the observed data is  $Y$  and there is a vector of latent variables  $Z \in \mathcal{Z} \subseteq \mathfrak{R}^n$ . Following a landmark contribution by Tan, Tian and Ng (2003, 2009) we know that

$$p(\theta|Y) \propto \frac{p(\theta|Y, Z_o)}{p(Z_o|\theta, Y)}, \forall Z_o \in \mathcal{Z}. \quad (1)$$

A similar formula holds for the latent variables:

$$p(Z|Y) \propto \frac{p(Z|Y, \theta_o)}{p(\theta_o|Z, Y)}, \forall \theta_o \in \Theta. \quad (2)$$

Tan, Tian and Ng (2003, 2009) have proposed the following method of recovering the posterior distribution of the parameters known as **IBF (Inverse Bayes Formula)**. Since  $p(\theta|Y) \propto p(\theta|Y, Z_o)p(Z_o|Y)$  it is clear that if we can obtain a random sample, say  $\{z^{(j)}, j = 1, \dots, J\}$  from  $Z_o|Y$  and we can obtain another random sample,  $\{\theta^{(j)}, j = 1, \dots, J\} \sim \theta|Y, z^{(j)}, j = 1, \dots, J$ , then  $\{\theta^{(j)}, j = 1, \dots, J\}$  is actually a random sample from the posterior  $p(\theta|Y)$ . The major difficulty is to draw  $z^{(j)}$  from  $p(Z_o|\theta, Y)$ . Tan, Tian and Ng (2003, 2009) have proposed the following approach. First,  $Z^* = \{z^{(j)}, j = 1, \dots, J\}$  is drawn from  $z|Y, \theta_o$  for some  $\theta_o \in \Theta$ . The value of  $\theta_o$  should be close to the posterior mean or mode. Second,  $z^{(j)}$ s are drawn from  $Z^*$  without replacement and probabilities

$$w_j = \frac{1/p(\theta_o|Y, z^{(j)})}{\sum_{j=1}^J 1/p(\theta_o|Y, z^{(j)})}, j = 1, \dots, M, \quad (3)$$

where  $M < J$ .

Another alternative is to use

$$p(\theta|Y) \propto \frac{p(\theta|Y, z_o)}{p(z_o|Y, \theta)}, \forall \theta \in \Theta, \quad (4)$$

and for some  $z_o \in \mathcal{Z}$ . If we can draw random samples from the ‘‘complete data’’ posterior,  $p(\theta|Y, z_o)$ , which is often possible, and the random sample is, say,  $\{\theta^{(j)}, j = 1, \dots, J\}$ , then we can use sampling without replacement using probabilities

$$w_j = \frac{1/p(z_o|Y, \theta^{(j)})}{\sum_{j=1}^J 1/p(z_o|Y, \theta^{(j)})}, j = 1, \dots, M. \quad (5)$$

In this paper we consider a stochastic volatility model (Jacquier, Polson and Rossi, 1994) for which a number of Markov Chain Monte Carlo (MCMC) algorithm has been proposed.<sup>1</sup>

$$\begin{aligned} y_t &= e^{h_t/2}\xi_t, \xi_t \sim \text{i.i.d.}\mathcal{N}(0, 1), t = 1, \dots, n, \\ h_t &= \alpha + \rho h_{t-1} + \varepsilon_t, \varepsilon_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2), \\ h_0 &\sim \mathcal{N}\left(\frac{\alpha}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right). \end{aligned} \quad (6)$$

The stochastic volatility model is based on the stylized fact that returns are unpredictable but squared returns are autocorrelated. One of the first approaches to model this dependence has been the Autoregressive Conditional Heteroskedasticity (ARCH) model in which we assume:

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2, \alpha_0, \alpha_1 \geq 0. \quad (7)$$

The ARCH model predicts that the conditional variance depends on lagged squared returns. The Generalized

---

<sup>1</sup>See Kim, Shephard and Chib (1998), Pitt and Shephard (1999), Chib, Nardari and Shephard (2002). A less popular choice is simulated maximum likelihood (Danielsson, 1994).

ARCH or GARCH model assumes:

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 h_{t-1}, \quad \alpha_0, \alpha_1, \alpha_2 \geq 0, \quad (8)$$

and incorporates lagged conditional variances into the model. Of course, the model can be extended easily to the GARCH(p,q) model given by:

$$h_t = \alpha_0 + \sum_{l=1}^p \alpha_l y_{t-l}^2 + \sum_{l=1}^q \alpha_{p+l} h_{t-l}, \quad \alpha_0, \alpha_l \geq 0, \quad l = 1, \dots, p+q. \quad (9)$$

In practice, it has been found that *stochastic* volatility of the form  $h_t = \alpha + \rho h_{t-1} + \varepsilon_t$ ,  $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2)$  performs best but it is not used very often due to complexities associated with constructing the likelihood function. These difficulties were mitigated by using MCMC methods in Jacquier et al. (1994) and since then a large literature has emerged on implementing statistical inference in stochastic volatility models.

Suppose that the prior is  $p(\alpha, \rho, \sigma) \propto \sigma^{-1}$ . The complete data posterior is:

$$p(\theta, \mathbf{h}|Y) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n \left[ h_t + y_t^2 e^{-h_t} + \frac{(h_t - \alpha - \rho h_{t-1})^2}{\sigma^2} \right] \right\}, \quad (10)$$

where the parameter is  $\theta = [\alpha, \rho, \sigma]'$ ,  $h = [h_1, \dots, h_n]'$ , and  $Y = [y_1, \dots, y_n]'$ . The posterior of the parameters

$$p(\theta|Y) = \int_{\Theta} p(\theta, \mathbf{h}|Y) d\mathbf{h}, \quad (11)$$

cannot be computed in closed form as the integral is not available analytically, and, even worse, it cannot be expressed as a product of univariate integrals. Relative to the class of problems considered by Tan, Tian and Ng (2003, 2009), (11) is more difficult due to the presence of the multivariate integral. The only available techniques to perform posterior inference use Markov Chain Monte Carlo (MCMC) methods which can be difficult to converge due to inherent autocorrelation. Therefore, a non-iterative technique is needed. Standard MCMC techniques (Jacquier, Polson and Rossi, 1994) are based on Gibbs sampling using  $\theta|\mathbf{h}, Y$  which is trivial and  $h_t|h_{t-1}, h_{t+1}, \theta, Y$  which slows down considerably the Gibbs sampler.

Adapting (1) we have:

$$p(\theta|Y) \propto \frac{p(\theta|Y, \mathbf{h}^o)}{p(\mathbf{h}^o|\theta, Y)}, \quad \forall \mathbf{h}^o \in \mathfrak{R}^n. \quad (12)$$

Additionally, we have:

$$p(\mathbf{h}|\theta, Y) \propto \frac{p(\mathbf{h}|Y, \theta)}{p(\theta|\mathbf{h}, Y)}, \quad \forall \theta \in \Theta. \quad (13)$$

The posterior in (12) is low dimensional and therefore we can use numerical integration techniques to obtain marginal posterior densities of  $\alpha, \rho$  and  $\sigma$  or direct Monte Carlo:

$$\begin{aligned} [\alpha, \rho]'|\sigma, Y &\sim \mathcal{N}(\hat{\beta}, V), \\ \frac{\sum_{t=1}^n (h_t^o - \alpha - \rho h_{t-1}^o)^2}{\sigma^2} | \alpha, \rho, Y &\sim \chi_n^2, \end{aligned} \quad (14)$$

where  $\hat{\beta} = (X'X)^{-1}X'\mathbf{h}^o$ ,  $V = \sigma^2(X'X)^{-1}$ ,  $X = [1, h_{t-1}, t = 1, \dots, T]$ ,  $\mathbf{h}^o = [h_1^o, \dots, h_n^o]'$ .<sup>2</sup>

<sup>2</sup>For simplicity we drop terms related to  $\alpha, \rho$  and  $\sigma$  that come through the prior of  $h_0$  in (6). These terms are accommodated via rejection sampling.

## 2 Model and Bayesian procedures

The problem is to select appropriately the vector of volatilities  $\mathbf{h}^\circ$ . We can, for example, use *realized volatility*. An alternative<sup>3</sup> is to solve for  $h_t^*$  the following system of equations:

$$\begin{aligned} (1 - y_t^2 e^{-h_t}) + \frac{2}{\sigma^2} [(1 + \rho^2)h_t - \alpha(1 - \rho) - \rho(h_{t-1} + h_{t+1})] &= 0, \quad t = 1, \dots, T-1, \\ (1 - y_T^2 e^{-h_T}) + \frac{2}{\sigma^2} [h_T - \alpha - \rho h_{T-1}] &= 0, \quad t = 1, \dots, T-1, \end{aligned} \quad (15)$$

which comes from maximizing (10) with respect to  $\mathbf{h}$ . If we linearize the term  $e^{-h_t} \simeq 1 - h_t$ , the approximate solution is given as follows:

$$\begin{aligned} \left\{ y_t^2 + \frac{2(1+\rho^2)}{\sigma^2} \right\} h_t &= -(1 - y_t^2) + \frac{2}{\sigma^2} \{ \alpha(1 - \rho) - \rho(h_{t-1} + h_{t+1}) \}, \quad t = 1, \dots, T-1, \\ \left\{ y_T^2 + \frac{2(1+\rho^2)}{\sigma^2} \right\} h_T &= -(1 - y_T^2) + \frac{2}{\sigma^2} \{ \alpha + \rho h_{T-1} \}. \end{aligned} \quad (16)$$

The system in (15) can be solved iteratively.<sup>4</sup> starting from the solution of (16) which must also be obtained after a few iterations or just a single iteration away from  $\mathbf{h} = \mathbf{0}$ . The single iteration produces the estimates:

$$\begin{aligned} \left\{ y_t^2 + \frac{2(1+\rho^2)}{\sigma^2} \right\} \hat{h}_t &= -(1 - y_t^2) + \frac{2\alpha(1-\rho)}{\sigma^2}, \quad t = 1, \dots, T, \\ \left\{ y_T^2 + \frac{2(1+\rho^2)}{\sigma^2} \right\} \hat{h}_T &= -(1 - y_T^2) + \frac{2\alpha}{\sigma^2}, \end{aligned} \quad (17)$$

which can be used as starting values for the iterations in (15).

If we use (13) there seems to exist another way to proceed. The numerator of this expression is:

$$p(\mathbf{h}|Y, \theta) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n \left[ h_t + y_t^2 e^{-h_t} + \frac{(h_t - \alpha - \rho h_{t-1})^2}{\sigma^2} \right] \right\}, \quad (18)$$

and the denominator is

$$p(\theta|\mathbf{h}, Y) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^n (h_t - \alpha - \rho h_{t-1})^2 \right\}. \quad (19)$$

which holds for arbitrary  $\theta \in \Theta$ . Taking the ratio of the two we have from (13):

$$p(\mathbf{h}|Y) \propto \exp \left\{ -\frac{1}{2} \sum_{t=1}^n (h_t + y_t^2 e^{-h_t}) \right\}. \quad (20)$$

This is clearly unimodal with a mode at  $h_t^* = \log y_t^2$ . Effectively, we can just use this simple solution to obtain the values of  $\{h_t^\circ\}$  needed in (14). This approximation does not turn out to perform very well in the application that we consider in the next section, so we do not give it further consideration.

## 3 Application

We apply the new techniques to a series of daily exchange rates of British pound versus the US dollar (1/1/2011-12/19/2015) for a total of 1,811 observations. The results are reported in Table 1 and Figure 1. MCMC results were obtained from a Gibbs sampler (see Chib et al., 2002) using 15,000 iterations discarding the first 5,000 to mitigate possible start up effects. Results from MCMC and the non-iterative methods (both exact and linearized) are quite close suggesting that accurate inferences can be produced at a fraction of the cost required for MCMC.

<sup>3</sup>The realized volatility was not found to work well in artificial data so we do not recommend it in practice.

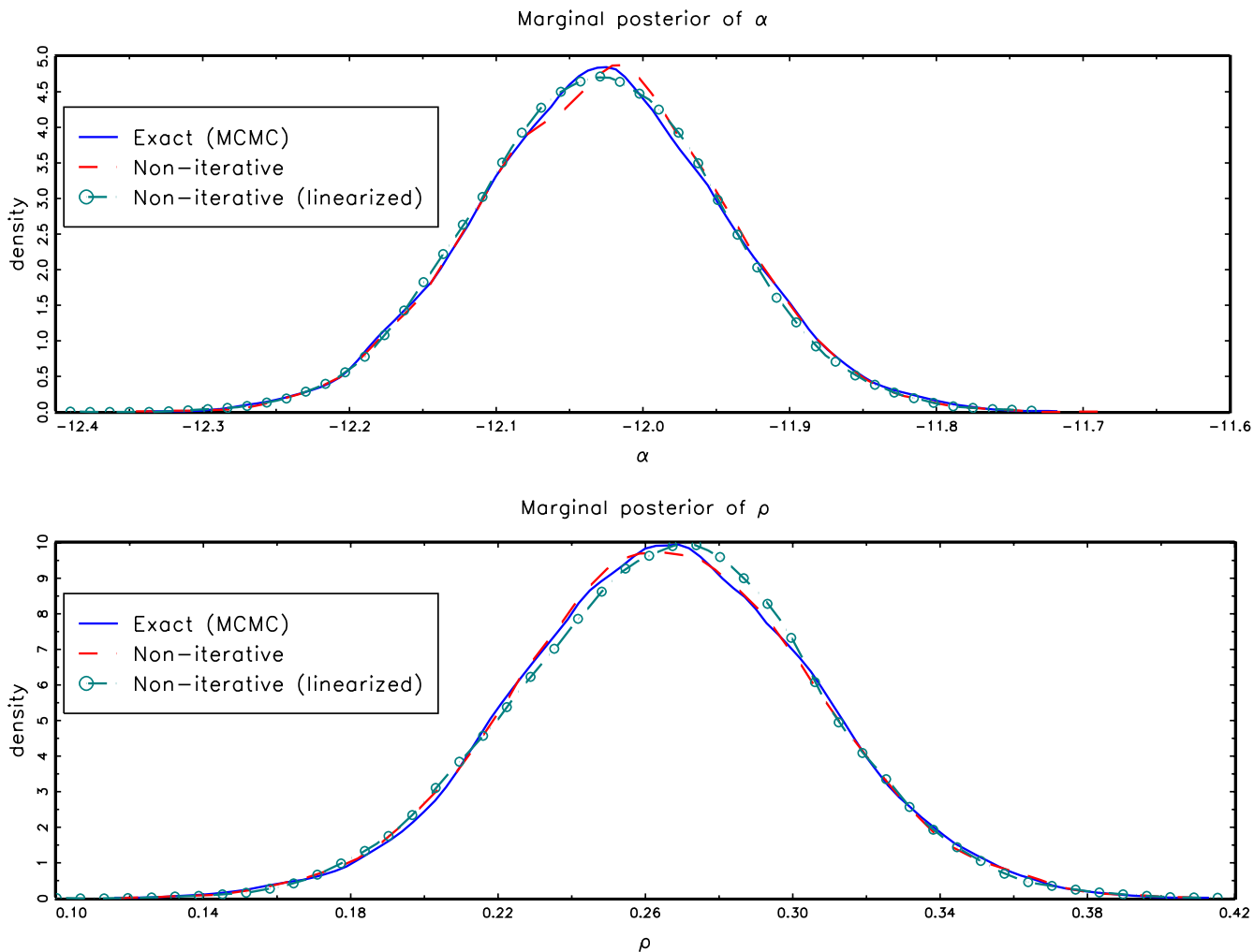
<sup>4</sup>The iterative solution is not necessary since the system is sparse. Using sparse solvers is possible but we did not implement the computation in this paper. Apparently, by non-iterative in the title of the paper we mean that the approach does not use MCMC.

Table 1: Empirical results

parameter	OLS <sup>(a)</sup>	MCMC <sup>(b)</sup>	Non-iterative <sup>(b)</sup>	Non-iterative (linearized) <sup>(b)</sup>
$\alpha$	-11.491 (0.324)	-12.030 (0.0838)	-12.029 (0.0830)	-12.028 (0.0835)
$\rho$	0.107 (0.025)	0.2658 (0.0404)	0.2651 (0.0401)	0.2655 (0.0422)

Notes: (a) OLS estimates of  $\log y_t^2$  on a constant and  $\log y_{t-1}^2$ . (b) Posterior means and posterior standard deviations in parentheses. Monte Carlo errors were 0.0027 for the constant term and 0.0011 for parameter  $\rho$ .

Figure 1. Comparison of marginal posterior distributions



Although the approximation in (17) does not perform well it does provide excellent initial conditions for the iterative solution of (15) or the iterative solution of (16).

## 4 Monte Carlo evidence

In this section we consider the stochastic volatility model in (6) with  $\alpha = -0.7$ ,  $\rho = 0.9$  and  $\sigma = 0.1$ . The sample size is  $T = 2,000$  and, alternatively,  $T = 10,000$ . We compare our results with MCMC methods described in Chib

Table 2: Monte Carlo results

	$T = 2,000$				$T = 10,000$			
	Bias		RMSE		Bias		RMSE	
	MCMC	IBF	MCMC	IBF	MCMC	IBF	MCMC	IBF
$\alpha$	-0.0032	-0.0030	0.0081	0.0081	-0.00012	-0.00012	0.00045	0.00045
$\rho$	-0.0017	-0.0016	0.0030	0.0030	-0.000015	-0.000015	0.00021	0.00021
$\sigma$	0.0015	0.0015	0.0044	0.0045	0.00032	0.00032	0.00043	0.00043

Notes: MCMC denotes Markov Chain Monte Carlo implemented as in Chib, Nardari and Shephard (2006). IBF denotes results from the Inverse Bayes Formula, see (1) and (3).

et al. (2002). Their MCMC techniques are based on the fact that, based on (6), we have:

$$\log y_t^2 = h_t + e_t, \quad (21)$$

where  $e_t = \log \xi_t^2$ . The distribution of  $e_t$  can be approximated using a mixture-of-normals as described in Chib, Nardari and Shephard (2006). As  $h_t$  follows a first-order autoregressive model, MCMC can be implemented easily using the Kalman filter. To implement MCMC we use 15,000 iterations discarding the first 5,000 to mitigate possible start up effects. To implement our technique we use  $J = 2,500$  and  $M = 2000$ . Moreover, we do not use linearization as in the previous section although the results are quite close and it can be recommended in practice. We use 5,000 replications to provide Monte Carlo evidence, reported in Table 2.

From the results in Table 2 it turns out that the behavior of MCMC and IBF is quite similar although IBF can be implemented using less computing effort and much less computing time, especially when the sample size is large ( $T = 10,000$ ).

## Conclusions

We propose a new non-iterative, very simple, Bayesian inference procedures for the stochastic volatility model. Certain of our results turn out to be good only as initial conditions in solving a large linear system, in terms of volatilities. The only requirement of our approach is to solve this linear system. The marginal posterior distributions of the parameters can be computed easily using direct Monte Carlo yielding independent samples from the joint posterior. We illustrate the performance of our method in a time series of GBP-USD exchange rates. We believe that this approximation will be essential in stochastic volatility models with leverage (Omori et al, 2007) and multivariate stochastic volatility models (Harvey et al, 1994 and Chib et al., 2006). This line of research is currently under investigation by the author. In particular, our construction is relevant because multivariate stochastic volatility models can be reduced to a sequence of univariate stochastic volatility models (e.g. Lopes et al, 2012).

## References

- Chib, S., F. Nardari, N. Shephard (2002). Markov Chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, 281-316.
- Chib, S., F. Nardari, N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics* 134, 341-371.
- Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics* 61, 375-400.
- Jacquier, E., N. G. Polson, P. E. Rossi (1994). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics* 12, 371-389.

- Harvey, A., E. Ruiz, N. Shephard (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies* 61, 247-264.
- Kim, S., N. Shephard, S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* 65 (3), 361-393.
- Lopes, H.F., R.E. McCulloch, R.S.Tsay (2012). Cholesky Stochastic Volatility Models for High-Dimensional Time Series, working paper.
- Omori, Y., S. Chib, N. Shephard, J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140, 425-449.
- Pitt, M., Shephard, N. (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association* 94, 590-599
- Tan, M., G.-L. Tian, K.W. Ng (2003). A noniterative sampling method for computing posteriors in the structure of EM-type algorithms. *Statistica Sinica* 13, 625-639.
- Tan, M., G.-L. Tian, K.W. Ng (2003). Non-iterative sampling-based Bayesian methods for identifying change points in the sequence of cases of haemolytic uraemic syndrome. *Computational Statistics and Data Analysis* 53 (9), 3314-3323.