Statistics and Computing **15**: 103–111, 2005

Bayesian analysis of the unobserved ARCH model

STEFANOS G. GIAKOUMATOS*, PETROS DELLAPORTAS*
and DIMITRIS N. POLITIS**

**Department of Statistics, Athens University of Economics and Business, Patission 76, 10434, Athens, Greece*
petros@aueb.gr
***Department of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112, USA*

Received August 2002 and accepted October 2004

The Unobserved ARCH model is a good description of the phenomenon of changing volatility that is commonly appeared in the financial time series. We study this model adopting Bayesian inference via Markov Chain Monte Carlo (MCMC). In order to provide an easy to implement MCMC algorithm we adopt some suitable non-linear transformations of the parameter space such that the resulting MCMC algorithm is based only on Gibbs sampling steps. We illustrate our methodology with data from real world. The Unobserved ARCH is shown to be a good description of the exchange rate movements. Numerical comparisons between competing MCMC algorithms are also presented.

Keywords: auxiliary variables, ARCH components, Markov chain Monte Carlo, GARCH

1. Introduction

Many financial time series, such as stock returns and exchange rates, can be successfully modeled by assuming that the error variance fluctuates over time. Thus, such models can capture a usual phenomenon, often met in financial time series, the "volatility clustering". The familiar modeling approaches are the Autoregressive Conditional Heteroskedasticity. (ARCH) models and their variants (Bollerslev, Engle and Nelson 1994). An alternative to those models is given by the Stochastic Volatility models. For a recent good description of the stochastic volatility models see: Shephard (1996) and Ghysels, Harvey and Renault (1996).

We focus our attention on a member of a class of models, introduced by Harvey, Ruiz and Sentana (1992), the Unobserved ARCH model, first explicitly presented by Shephard (1996). In this model the ARCH component is observed with error, or, it may be seen as a latent process. In a recent paper, Fiorentini, Sentana and Shephard (2004) discuss extensively the need to study models in which an ARCH process is used as a latent process. They point out that such modelling mechanisms can tackle important empirical problems in broad areas which include aggregate models, bid/ask prices, target interest rates and future contracts. Fiorentini, Sentana and Shephard (2004) concentrate on a multivariate heteroskedasticity model which may be viewed as a factor-type multivariate extension of the model we propose here.

We adopt Bayesian inference and we propose an easy to implement and fast to converge MCMC algorithm. To achieve this, we transform the parameter space so that the resulting full conditional posterior densities have simplified forms. Moreover we adopt the Auxiliary Variable (AV) sampler (Swendsen and Wang 1987) and we include a set of latent variables in the full posterior density of the model such that all the full conditional densities to be of known forms. The MCMC algorithm we propose consists of *only* Gibbs steps. This emanated from the desire to develop an algorithm which is straightforward to use (unlike Metropolis-Hastings, Gibbs sampling requires no tuning) and has better convergence behaviour than existing MCMC samplers. Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998) also used MCMC techniques to study the stochastic volatility model, but their algorithms required Metropolis steps.

The remainder of the paper is organized as follows. In Section 2 we present in some detail the Unobserved ARCH model and some of its theoretical properties. In Section 3 we describe the non-linear transformations of the parameter space of the model that we adopt and in Section 4 we propose our AV algorithm. In Section 5 we illustrate this algorithm with real data and in

0960-3174

Section 6 we compare the proposed algorithm with other existing popular algorithms. In Section 7 we discuss the overall fit of the model for our data.

2. The unobserved ARCH model

The Unobserved ARCH model is presented by Shephard (1996). The ARCH components in this model are observed with errors. The form of this model can be written using the following hierarchical structure of conditional densities:

$$y_t|f_t,\sigma^2 \sim \mathbf{N}(f_t,\sigma^2); \quad f_t|f_{t-1},a,b,f_0 \sim \mathbf{N}(0,h_t);$$
$$h_t = a + b\cdot f_{t-1}^2; \quad (1)$$

where $y_1,\ldots,y_T$ is a realization of the process, f_t is the Unobserved ARCH component at time t, f_0 is the initial state or the "history" of the unobserved components and $\mathbf{N}(\cdot,\cdot)$ is the Normal distribution. To obtain $h_t > 0$, the parameters a and b are restricted to be positive. The additional restriction $0 < b \leq 1$ is placed so that the ARCH component of the model to be covariance stationary (Engle 1982).

The restriction $b > 0$ is imposed to this model so that the parameters are identifiable. Specifically, when $b = 0$, y_t is a sum of two independent Gaussian white noises with variances σ^2 and a respectively, which cannot be separately identified on the basis of the sample information alone.

Note that the unobserved component f_t is not measurable with respect to the available information at time t, something which characterizes this class of models. The unconditional and conditional variances of y_t are given by $Var\,(y_t) = \sigma^2 + a/(1-b)$ and $Var(y_t|y_{t-1},a,b) = \sigma^2 + h_t$. Therefore, the stochastic process y_t can be considered to have an underline variance on which it is added the variability which is caused by the effect of volatility clustering.

There are various reasons one might choose to use an Unobserved ARCH model instead of other standard models such as an observable GARCH. For example, note that if y_t follow an Unobserved ARCH(p) model as in (1), then it is easily shown that y_t^2 follow a non-Gaussian ARMA(p,p) process, mimicking the well-known nice property of the square returns of GARCH models. To see this, note that $y_t^2 = y_t^2 - \sigma^2 + \sigma^2 = f_t^2 + k_t + \sigma^2$ with $k_t = \sigma^2 e_t^2 + 2 f_t \sigma e_t - \sigma^2$. Now f_t^2 follows an AR(p) process since it can be written as $f_t^2 = a + \sum_{i=1}^p b_i \cdot f_{t-i}^2 + z_t$ with $z_t = f_t^2 - h_t$. Therefore, $z_t + a = b(B)\,f_t^2$, where $b(\cdot)$ is the pth degree polynomial ($b(\xi) = 1 - b_1\xi - \cdots - b_p\xi^p$) and B is the backward shift operator. Then, $b(B)\,y_t^2 = b(B)\,f_t^2 + b(B)\,k_t + (1-\sum_{i=1}^p b_i)\sigma^2 = (1-\sum_{i=1}^p b_i)\sigma^2 + a + w_t$ where $w_t = b(B)k_t + z_t$, and y_t^2 is an $ARMA(p,p)$ process since w_t is an $MA(p)$ process as a sum of an $MA(p)$ process and a white noise. Moreover, in a series of empirical studies in Giakoumatos (2004) that are not reported here, there was evidence that the predictive ability of the Unobserved ARCH model is better for one-step ahead predictions than GARCH(1,1) models.

3. Bayesian inference

The posterior density of the parameters of the Unobserved ARCH model can be extracted via Bayes theorem, by

$$[a,b,\sigma^2,f_0,\mathbf{f}\,|\,\mathbf{y}] \propto \prod_{t=1}^{T}([y_t\,|\,f_t,\sigma^2][f_t\,|\,f_{t-1},a,b]) \times [a,b,\sigma^2,f_0]. \quad (2)$$

(Throughout the paper the usual square bracket notation is used for joint, conditional and marginal densities.) The first two terms in the above product are derived from the hierarchical structure in (1) and the last term, $[a,b,\sigma^2,f_0]$, is the joint prior density of a, b, σ^2 and f_0. These parameters are assumed a priori independent and we choose improper priors for the a, b, σ^2 and a vague Normal density $\mathbf{N}(0,v)$ for f_0, so that the joint prior density takes the form $[a,b,\sigma^2,f_0] \propto (a\cdot\sigma^2)^{-1}\exp\{-0.5 f_0^2/v\}$. Note that the choice of priors is important since other functional forms of priors on a and b would destroy the algebraic mechanism which is essential for the derivation of the results of Section 4. However, an inverted gamma prior could be used for σ^2 without any serious complications. Using the above joint prior density, the joint posterior density (2) takes the form

$$[a,b,\sigma^2,f_0,\mathbf{f}\,|\,\mathbf{y}] \propto \frac{1}{\prod_{t=1}^T\sqrt{a+bf_{t-1}^2}}\exp\left\{-\frac{1}{2}\sum_{t=1}^T\left(\frac{f_t^2}{a+bf_{t-1}^2}\right)\right\} \times \frac{1}{a\sigma^{2\frac{T+2}{2}}}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^T(y_t-f_t)^2+\frac{f_0^2}{v}\right)\right\}. \quad (3)$$

The above posterior (3) is heavily parameterized and the full posterior conditional densities (i.e. the full posterior conditional density means, the posterior density of one parameter condition on all the remaining parameters) are not of standard forms. Therefore, the construction of the MCMC algorithm is not at all simple. In order to handle this problem, we adopt some *non-linear* transformations of the parameter space. Firstly, note that

Remark 1. If in the posterior density $\left[a,b,\sigma^2,f_0,\mathbf{f}\,|\,\mathbf{y}\right]$ defined in (3) we perform the following transformations $g=\sqrt{a/b}$ and $w_t = \sqrt{b/a}\,f_t$; for $t = 0,\ldots T$, the posterior density takes the form

$$\left[g,b,\sigma^2,\mathbf{w}\,|\,\mathbf{y}\right] \propto \frac{1}{\prod_{t=1}^T\sqrt{1+w_{t-1}^2}}\exp\left\{-\frac{1}{2b}\sum_{t=1}^T\frac{w_t^2}{(1+w_{t-1}^2)}\right\} \times \frac{1}{\sigma^{2\frac{T+2}{2}}b^{\frac{T}{2}}}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^T(y_t-gw_t)^2+\frac{(gw_0)^2}{v}\right)\right\}, \quad (4)$$

where $\mathbf{w} = (w_0,\ldots,w_T)$.

P1: xxx
SJNW132-04-5386202 March 4, 2005 21:5 Char Count=

Proof: Note that the Jacobian of the above transformations is $|J| = 2bg^{T+2}$. The remaining calculations are straightforward. □

By using Remark 1, the resulting posterior density (4) has full conditional densities of a rather convenient form. In particular,

- $$\left[\sigma^2|\cdot\right] \equiv \mathbf{IG}\left(\frac{T}{2}, \frac{1}{2}\sum_{t=1}^{T}(y_t - gw_t)^2\right),$$
  where $\mathbf{IG}(a, b)$ denotes the Inverse Gamma density with mean $b/(a-1)$; the notation $|\cdot$ implies conditioning on all the remaining parameters.
- $$[b|\cdot] \equiv \mathbf{IG}\left(\frac{T-2}{2}, \frac{1}{2}\sum_{t=1}^{T}\frac{w_t^2}{1+w_{t-1}^2}\right)\mathbf{I}(b \le 1),$$
  where $\mathbf{I}(\cdot)$ is the indicator function.
- $$[g|\cdot] \equiv \mathbf{N}(m,s)\,\mathbf{I}(g \ge 0),$$
  where
  $$m = \left(v\sum_{t=1}^{T} w_t y_t\right) \Big/ \left(\sigma^2 w_0^2 + v\sum_{t=1}^{T} w_t^2\right)$$
  and
  $$s = (\sigma^2 v) \Big/ \left(\sigma^2 w_0^2 + v\sum_{t=1}^{T} w_t^2\right).$$
- $$[w_0|\cdot] \propto \mathbf{ND}\left(0, \frac{v}{g^2}\right)\frac{1}{\sqrt{1+w_0^2}}\exp\left\{-\frac{1}{2b}\frac{w_1^2}{(1+w_0^2)}\right\}.$$
- $$[w_t|\cdot] \propto \mathbf{ND}\left(m_t, s_t^2\right)\frac{1}{\sqrt{1+w_t^2}}\exp\left\{-\frac{1}{2b}\frac{w_{t+1}^2}{(1+w_t^2)}\right\},$$
  for $t = 1, \ldots, T-1$.
- $$[w_T|\cdot] \equiv \mathbf{N}\left(m_T, s_T^2\right),$$

where $\mathbf{ND}(\cdot,\cdot)$ denotes the p.d.f. of the Normal distribution and m_t and s_t^2 are given by

$$m_t = \frac{y_t g b(1+w_{t-1}^2)}{gb(1+w_{t-1}^2)+\sigma^2}, \quad s_t^2 = \frac{\sigma^2 b(1+w_{t-1}^2)}{g^2 b(1+w_{t-1}^2)+\sigma^2}. \tag{5}$$

Again, the full conditional densities of w_t, for $t = 0, \ldots, T-1$, are not of known forms. One way to deal with it, is to use Metropolis-Hastings steps (Hastings 1970, Metropolis *et al.* 1953) which allow us to sample from non-standard densities. We tried a random walk Metropolis-Hastings step with a Normal proposal density with variance given by s_t^2. For a series of data sets we have analyzed, the probability of acceptance is approximately 0.5, a value which has been considered satisfactory by Chib and Greenberg (1995a).

However, note that the full conditional densities of w_t, $t = 0, \ldots, T-1$, can be written as

$$[w_t|\cdot] \propto \mathbf{N}\left(m_t, s_t^2\right)\Psi(w_t), \quad t = 0, \ldots, T-1$$

where $\Psi(w_t)$ is a function of w_t. In that case, we can follow Chib and Greenberg (1994) (see also Chib and Greenberg 1995b) and sample from these full conditional densities by a Metropolis-Hastings step using as proposal density $\mathbf{N}(m_t, s_t^2)$. In this case the probability of acceptance reduces to $\min\{1, \Psi(w_t')/\Psi(w_t)\}$, where w_t' is the proposal value.

Another way to sample from w_t, $t = 0, \ldots, T-1$, is to use *AV* sampling techniques (Swendsen and Wang 1987, Edwards and Sokal 1988, Besag and Green 1993, Hidgon 1998, Damien, Wakefield and Walker 1999, Neal 2003). The way this is achieved becomes evident in the next section.

4. The auxiliary variable sampling

The basic idea of *AV* sampling is that the parameter space of a posterior density can be increased by including extra latent positive variables which make the resulting posterior density more tractable by sampling methods. Apart from the simplicity, the *AV* sampling has many other promising properties fully examined by Damien *et al.* (1999), Mira and Tierney (2002) and Roberts and Rosenthal (1999).

We use *AV* sampling to construct an MCMC algorithm from which we easily sample from the posterior of the Unobserved ARCH model. We expand the parameter space, introducing $2T$ auxiliary variables such that the resulting MCMC algorithm to consists of only Gibbs steps. These auxiliary variables do not have a useful interpretation but they are simply used as means to facilitate the implementation of the MCMC algorithm. To construct our proposed algorithm we use the following Remark.

Remark 2. If we include $2T$ positive latent variables $\mathbf{u} = (u_1, \ldots, u_T)$ and $\mathbf{k} = (k_1, \ldots, k_T)$ in the posterior density (4) such that the resulting joint density is given by

$$\begin{aligned}
&[g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k}|\mathbf{y}]\\
&\propto \frac{1}{\sigma^{2\frac{T+2}{2}} b^{\frac{T}{2}}}\left(\prod_{t=1}^{T}\mathbf{I}\left(u_t \le \frac{1}{\sqrt{1+w_{t-1}^2}}\right)\right)\\
&\quad\times\left(\prod_{t=1}^{T}\mathbf{I}\left(k_t \le \exp\left(-\frac{1}{2b}\frac{w_t^2}{(1+w_{t-1}^2)}\right)\right)\right)\\
&\quad\times\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{t=1}^{T}(y_t - gw_t)^2 + \frac{(gw_0)^2}{v}\right)\right\},
\end{aligned} \tag{6}$$

then, the marginal density $[g, b, \sigma^2, \mathbf{w}\,|\,\mathbf{y}]$ is given by (4).

P1: xxx
SJNW132-04-5386202 March 4, 2005 21:5 Char Count=

The above Remark guarantees that a MCMC algorithm which obtains samples from $[\mathbf{u}, \mathbf{k}, g, b, \mathbf{w}, \sigma^2 \,|\, \mathbf{y}]$ obtains also samples from $[g, b, \mathbf{w}, \sigma^2 \,|\, \mathbf{y}]$. To utilize Remark 2, we need to further elaborate on the resulting full conditional densities of $[g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k}|\mathbf{y}]$. In fact, it is readily evident that it is more convenient to use some forms of conditional densities appropriately marginalised over some parameters (Chib and Carlin 1999). In particular, to sample from $[g, b, \sigma^2, \mathbf{w}, \mathbf{u}, \mathbf{k}|\mathbf{y}]$ we use the full conditional densities of g and σ^2, which are presented in Section 2, because they are independent of $\mathbf{u}$ and $\mathbf{k}$. For the remaining of the parameters, we use

- Instead of sampling from $[b|\cdot] \equiv [b \,|\, \mathbf{k}, \mathbf{w}]$ we sample from

$$[b|\mathbf{w}] \equiv \mathbf{IG}\left(\frac{T-2}{2}, \frac{1}{2}\sum_{t=1}^{T}\frac{w_t^2}{1+w_{t-1}^2}\right)\mathbf{I}\,(b \le 1).$$

-

$$[u_t|\cdot] \equiv \mathbf{U}\left(0, \frac{1}{\sqrt{1+w_{t-1}^2}}\right),$$

  for all $t = 1, \ldots, T$.

-

$$[k_t|\cdot] \equiv \mathbf{U}\left(0, \exp\left\{-\frac{w_t^2}{2b\,(1+w_{t-1}^2)}\right\}\right),$$

  for all $t = 1, \ldots, T$.

-

$$[w_0|\cdot] \equiv \mathbf{N}\,(0, v)\,\mathbf{I}\left(u_1 \le \frac{1}{\sqrt{1+w_0^2}}\right) \cdot \mathbf{I}\left(k_1 \le \exp\left\{-\frac{w_1^2}{2b\,(1+w_0^2)}\right\}\right).$$

- Instead of sampling from

$$[w_t|\cdot] \equiv \left[w_t|u_{t+1}, k_t, k_{t+1}, g, b, w_{t-1}, w_{t+1}, \sigma^2, \mathbf{y}\right];$$

  for $t = 1, .., T-1$,
  we sample from

$$\begin{aligned}[w_t|u_{t+1}, k_{t+1}, g, b, w_{t-1}, w_{t+1}, \sigma^2, \mathbf{y}] &\equiv \mathbf{N}\left(m_t, s_t^2\right)\mathbf{I}\left(u_{t+1} \le \frac{1}{\sqrt{1+w_t^2}}\right)\\ &\times \mathbf{I}\left(k_{t+1} \le \exp\left\{-\frac{w_{t+1}^2}{2b(1+w_t^2)}\right\}\right),\end{aligned}$$

  where m_t and s_t^2 are defined in (5).
- Instead of sampling from

$$[w_T|\cdot] \equiv [w_T|k_T, g, b, w_{T-1}, \sigma^2, \mathbf{y}],$$

  we sample from

$$[w_T|g, b, w_{T-1}, \sigma^2, \mathbf{y}] \equiv \mathbf{N}\big(m_T, \mathbf{s}_T^2\big),$$

  where m_T and s_T^2 are defined in (5).

In order to sample from the above truncated Normal density we chose to use rejection sampling (Gelfand, Smith and Lee 1992) which resulted in our application of Section 5 to acceptance rate of 0.15. Clearly, the choice of two auxiliary variables for sampling $[w_t|\cdot]$ instead of one was motivated by the need to increase to increase this acceptance rate, since one auxiliary variable (say λ_t) would requires draws from Normal densities truncated on the region specified by

$$\mathbf{I}\left(\lambda_t \le \frac{1}{\sqrt{1+w_{t-1}^2}}\exp\left(-\frac{1}{2b}\frac{w_t^2}{\left(1+w_{t-1}^2\right)}\right)\right).$$

For the full conditional density of b, which is truncated Inverse Gamma density, we use the *AV* sampler, introducing a latent variable; see Appendix.

5. Application

We focus on the daily exchange rate of the Germany Marc (*DEM*) with respect to the Greek Drachma. To elaborate, let c_t be the exchange rate of a currency with respect to the drachma on day t; then data series is given by $y_t = \log(c_t/c_{t-1}) \cdot 100$, that represents the daily relative (percentage) change of the exchange rate since $\log\,(c_t/c_{t-1}) \simeq (c_t/c_{t-1}) - 1 = (c_t - c_{t-1})/c_{t-1}$ for $(c_t/c_{t-1}) \simeq 1$. Our data set (Fig. 1(a)) consists of 844 observations taken in the period 16/12/93–2/5/97. Using our proposed algorithm of Section 3, we obtain a sample from the posterior density of the parameters of the Unobserved ARCH model. Dropping the first 40000 iterations as burn-in, we sample 1 value every 200 iterations such that the final sample, which consist of 2000 values, to be approximately an independent and identically distributed sample from the marginal densities of the parameters of interest. Note that, this thinning of the sampling draws is only useful to save disk and memory space. This sample can be used to find the summary statistics and to plot the histograms of the posterior densities of the parameters a, b and σ^2. The posterior mean of b (0.969) indicates that the ARCH component of the DEM against Drachma is volatility persistent.

For the unobserved components f_t; $t = 1, \ldots, T$, we plot the time series of their posterior means in Fig. 1(b). When comparing this plot with the plot of the daily *DEM*/*Drachma* exchange rates (Fig. 1(a)), we notice that the Unobserved ARCH components model very well the volatility clustering of the observed data.

6. Comparison of different MCMC algorithms

Using the first 3000 iterations we can compare the *AV* sampling MCMC algorithm with respect to the Metropolis-Hastings algorithm (*MH*) and the Chib and Greenberg algorithm (*CG*). We

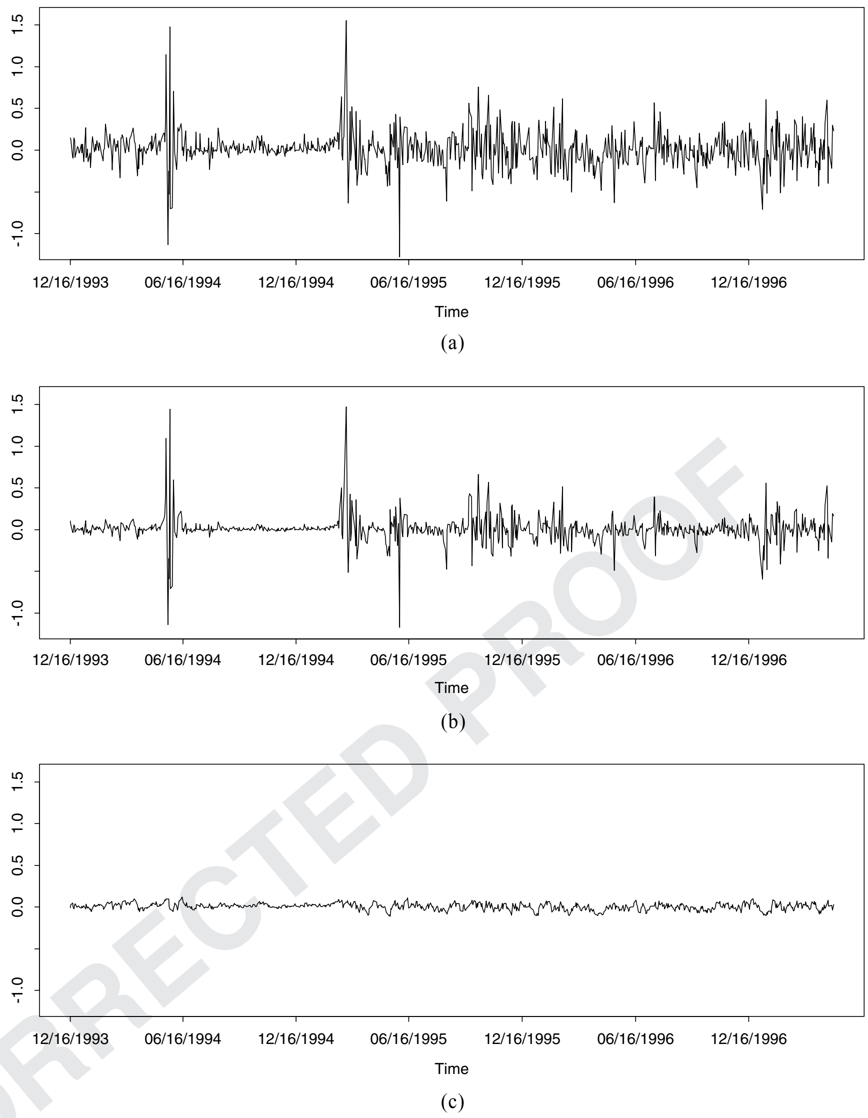

Fig. 1. *(a) DEM/Drachma exchange rates. (b) Posterior means of the Unobserved ARCH components. (c) Means of the smoothed Unobserved ARCH components.*

initialize the three algorithms from the same dispersed initial values. The results are presented in Figs. 2(a)–(c). It seems that for all the marginal densities *AV* and *CG* algorithms reach the target distribution much faster than MH. In particular, *AV* has the best rate of convergence in Figs. 2(a) and (b) (parameters a and b) and *CG* achieves the fastest convergence in Fig. 2(c) (parameter σ^2). To judge the overall convergence of the joint posterior density of (a, b, σ^2) we use the subsampling diagnostic that recently presented by Giakoumatos, Vrontos, Dellaportas and Politis (1999) which gauges at what point the chain "forgets" its starting points. This diagnostic uses as critical point that a chain gets in the target distribution the time that the coefficient of determination of a regression crosses a threshold d. The regression is based on the "range" of the $(1-a)100\%$ confidence region for the t quantile of the distribution of interest and on $1/\sqrt{N}$. The $(1-a)100\%$ confidence regions for the t quantile are calculated using subsampling techniques for subsequent portions of the chain and N is the number of iterations that are used to calculate the confidence regions. In our example we set $a = 0.05$, $t = 0.90$, $d = 0.99$, and we use the first 50000 iterations of the three algorithms. The results of the subsampling diagnostic are represented in Fig. 3. The *AV* algorithm needs approximately 13500 iterations to get in the target distribution when the CG algorithm needs approximately 15500 iterations and the MH approximately 22000 iterations. Note that the subsampling diagnostic is considered by its authors very 'conservative'.

Another usual visual inspection which characterizes the behaviour is based on the correlograms of the Markov chain. In Fig. 4 it seems that the autocorrelation of *AV* dies out much faster than the *CG* and *MH*.

The computational time for one iteration of the *MH* and *CG* algorithms is much higher than the *AV* algorithm. In general, the

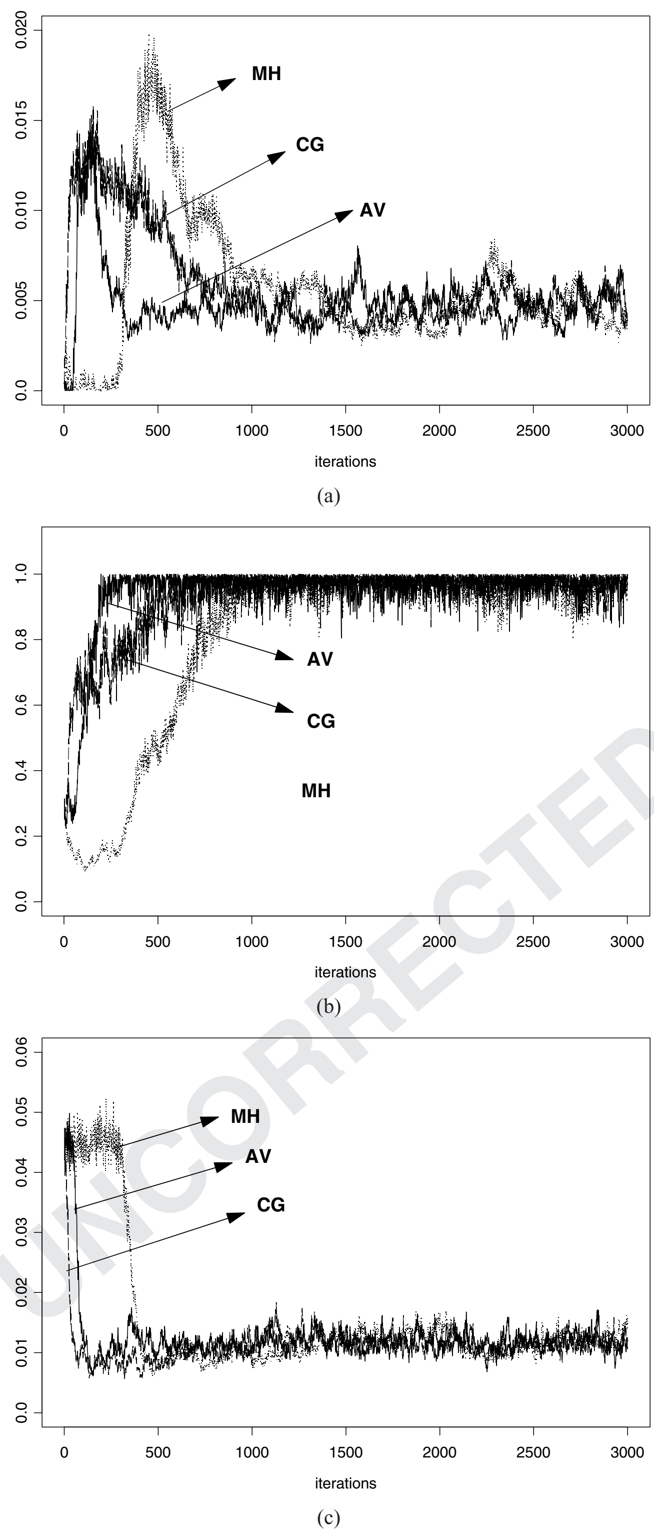

Fig. 2. *(a) Markov chain behavior for the three samplers for the parameter a. Solid line: AV sampler. Dotted line: MH sampler. Dashed line: CG sampler. (b) Markov chain behavior for the three samplers for the parameter b. Solid line: AV sampler. Dotted line: MH sampler. Dashed line: CG sampler. (c) Markov chain behavior for the three samplers for the parameter σ^2. Solid line: AV sampler. Dotted line: MH sampler. Dashed line: CG sampler.*

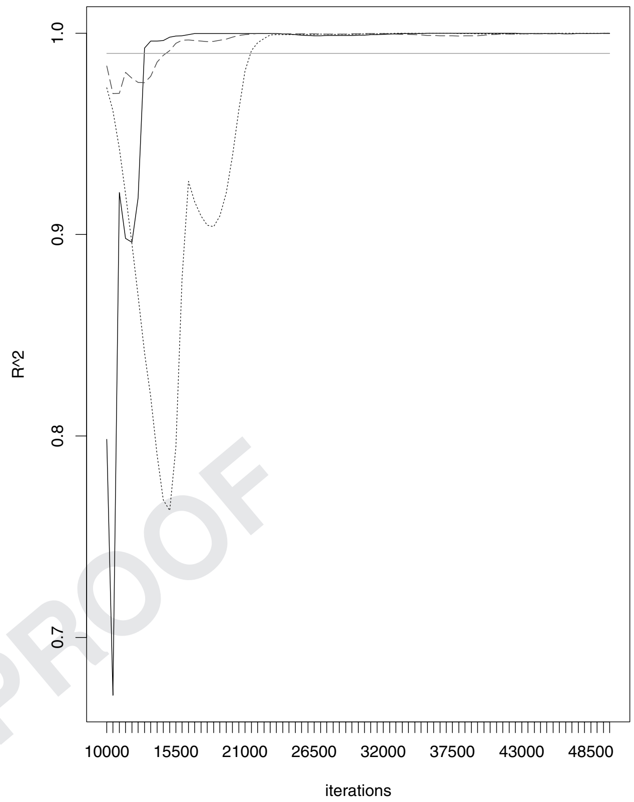

Fig. 3. *Coefficient of determination versus iterations. Solid line: AV sampler. Dotted line: MH sampler. Dashed line CG sampler. Straight line: threshold $d = 0.99$.*

expected computational times of *MH* and *CG* are more than 2 times greater than the expected time of *AV* algorithm.

Finally, we use the effective sample size (*ESS*) to compare the three algorithms (Neal 1993, Brooks 1999). The *ESS* is defined as the ratio n/τ—where n is the size of the sample, τ is the autocorrelation time calculated by $\tau = \sum_{-\infty}^{+\infty} \rho(t)$ and $\rho(t)$ is the autocorrelation function at lag t. It expresses, in terms of the variance of the sample mean, the number of independent observations which correspond to a sequence of n dependent observations. Using the first 3000 iterations of each of the three algorithms we estimate the *ESS* for (a, b, σ^2) and we keep the smallest of these as the minimum *ESS* of the corresponding algorithm. Moreover dividing the minimum *ESS* of each algorithm with the computational time that was needed for the 3000 iterations we get the minimum *ESS* per second. The minimum *ESS* and the minimum i per second are found to be (19.5, 0.1336), (5.1, 0.0148) and (4.6, 0.0135) for AV, CG and MH respectively. Again, it seems that the *AV* algorithm performs better than the other two algorithms for this data set.

In conclusion, it seems that the AV performs relatively better on the measures we considered. Both the *ESS* and the correlograms show a relative advantage whereas the subsampling diagnostic indicates that the difference with the *CG* algorithm is small. Although one might wonder how good these algorithms are in absolute, rather than in relative terms, we feel that AV

P1: xxx
SJNW132-04-5386202 March 4, 2005 21:5 Char Count=

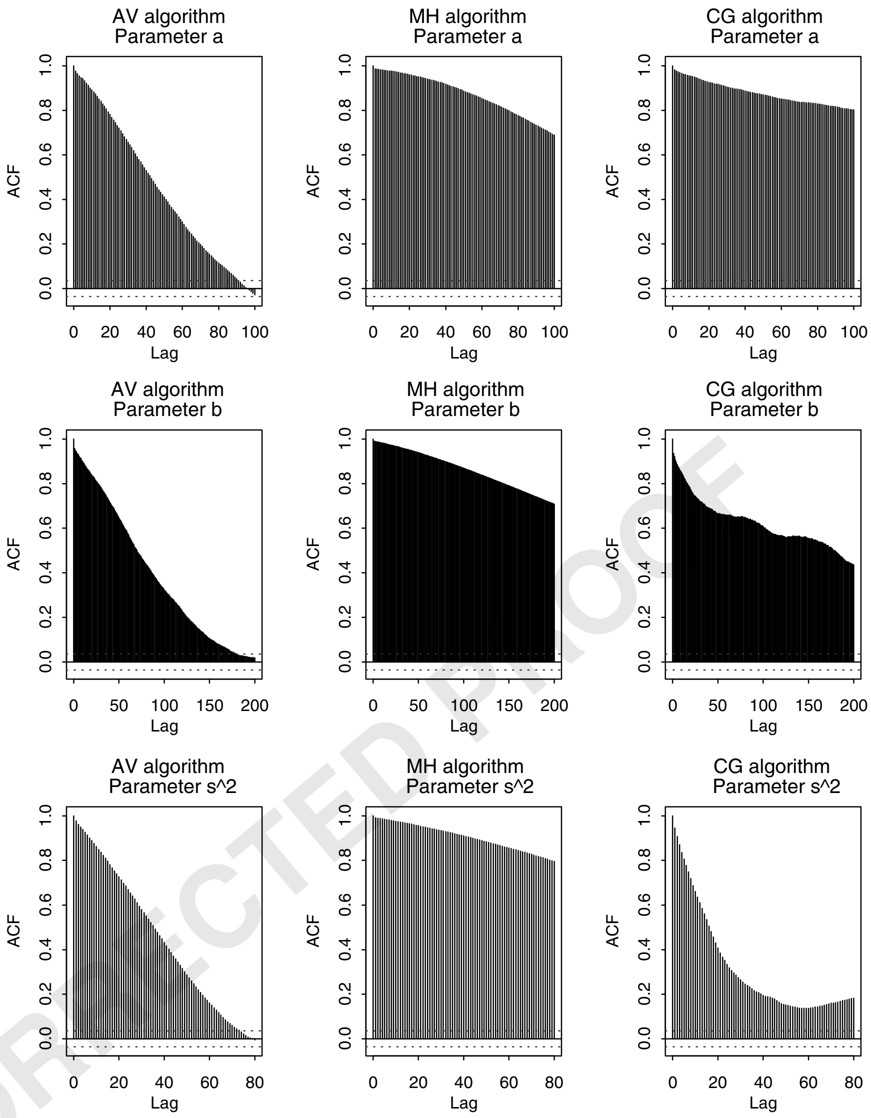

Fig. 4. *Autocorrelation function plots for the parameters of the model.*

algorithm performs generally well due to its implementation ease.

Another algorithm that might be worth investigating as an alternative to our *AV* sampler is the single-step Metropolis suggested by Shephard (1996) and Pitt (1997). Unfortunately, our posterior density produces a full conditional density of w_t that is a product of a normal density and a non-log-concave function. Therefore, the nice property of log-concavity that was utilized in Shephard (1996) does not exist in our model and we did not pursue this algorithm further.

7. Overall fit of the model

In order to check the overall fit of the model we follow Pitt and Shephard (1999) and we calculate the following residuals:

(1) Log-likelihood l_t; for $t = 1, \ldots, T$ where

$$l_{t+1} = \log([Z_{t+1} = y_{t+1} | \mathbf{\Psi}_t, \bar{\boldsymbol{\theta}}]),$$

y_{t+1} is the realization of the stochastic process at time $t+1$, $\mathbf{\Psi}_t$ is the available information up to time t and $\bar{\boldsymbol{\theta}}$ is the vector of the posterior means of the parameters of the Unobserved ARCH model.

As Pitt and Shephard (1999), we can estimate $[Z_{t+1} = \mathbf{y}_{t+1} | \Psi_t, \bar{\boldsymbol{\theta}}; M_k]$ via filtering methods. First we obtain samples of size K (in the applications of section 4 we use $K = 10000$) of the unobserved component f_t of the model, for $t = 1, \ldots, T$, using $f^i_{t+1} \sim [f_{t+1} | \mathbf{\Psi}_t, \bar{\boldsymbol{\theta}}]$, for $i = 1, \ldots, K$. The above density is easily derived for model (1). Then an estimate is given by

$$[\hat{Z}_{t+1} = y_{t+1} | \mathbf{\Psi}_t, \bar{\boldsymbol{\theta}}] = \frac{1}{K} \sum_{i=1}^{K} \left[\hat{Z}_{t+1} = y_{t+1} | f^i_{t+1}, \bar{\boldsymbol{\theta}} \right], \quad (7)$$

which is the result of the Monte Carlo integration of

$$[y_{t+1}|\boldsymbol{\Psi}_t, \bar{\boldsymbol{\theta}}] = \int [y_{t+1}|f_{t+1}, \bar{\boldsymbol{\theta}}][f_{t+1}|\boldsymbol{\Psi}_t, \bar{\boldsymbol{\theta}}]df_{t+1}.$$

This technique presupposes that we can evaluate and simulate from the density $[y_t|f_t, \bar{\boldsymbol{\theta}}]$ something which is true for the Unobserved ARCH model. For more details about this technique and the filtering method see Pitt and Shephard (1999).

(2) Standardized log-likelihood l_t^s, where

$$l_t^s = \frac{l_t - \mu_t^l}{\sigma_t^l},$$

where μ_t^l and σ_t^l are the mean and standard deviation of l_t, constructed as follow. We sample $y_{t+1}^{*i} \sim [y_{t+1}|\boldsymbol{\Psi}_t, \bar{\boldsymbol{\theta}}]$, for $i = 1, \ldots, S$ ($S = 100$ in our application). For each y_{t+1}^{*i} we evaluate l_{t+1}^i using the above methodology, and we calculate μ_t^l and σ_t^l as the sample mean and standard deviation of l_t^i.

The statistic l_t^s should have mean 0 and standard deviation 1 if the model and the parameters are correct. Large negative vales of l_t^s imply that an observation is less likely than the model would expect; and vise-versa for the positive values. Moreover we expect that $\sum_{t=1}^{T} l_t^s / T \to 0$ as $T \to \infty$ (under the weak law of large numbers).

In our case 26 l_t^s out from 844 ($26/844 \simeq 0.03$) are below -3 and 2 out of 844 are below -10. The mean of the l_t^s is about 0.0273 and the variance is approximately 2.14. Therefore the mean is close to zero (not statistically significant) but the variance is much larger than 1, indicating that there are a lot of very unlikely or very likely observations than expected but less in between!

Moreover, from the filtering procedure we easily calculate the smoothed Unobserved factors $\tilde{f}_t$, for $t = 1, \ldots, T$, using as an estimate the mean $\tilde{f}_t = K^{-1}\sum_{i=1}^{K}[f_t|\boldsymbol{\Psi}_t, \bar{\boldsymbol{\theta}}]$. These estimates are presented in Fig. 1(c).

Figure 5 presents the correlograms and the partial correlograms of the squares of the exchange rates and the squares of the posterior means of the Unobserved components f_t. It is remarkable how the squares of the Unobserved components mimic the behaviour of the squares of the observed data. Moreover the correlogram of y_t^2 indicates that this process may have an ARMA(1,1) behaviour. From the correlogram of f_t^2 it is clear that this behaviour is captured by the Unobserved components even thought they are based on ARCH(1).

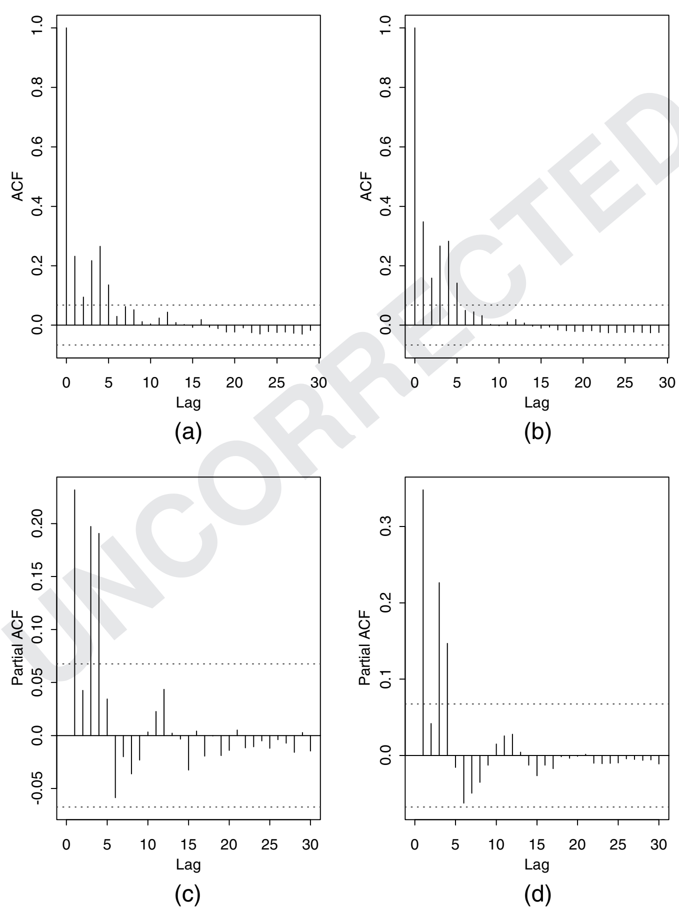

Fig. 5. *Correlograms and partial correlograms of the squares of the observed data and the squares of the posterior means of the Unobserved factor respectively. (a) Correlogram of the squares of the observed data. (b) Correlogram of the squares of the posterior means of the Unobserved factor. (c) Partial correlogram of the squares of the observed data. (d) Partial correlogram of the squares of the posterior means of the factor.*

Appendix

`Sampling from truncated Inverse Gamma`

We want to sample a value x from the density $[x] \propto \frac{1}{x^{a+1}}\exp(-\frac{b}{x})\mathbf{I}(x < d)$. Instead of sampling x, we sample y from $[y] \propto y^{a-1}\exp(-by)\mathbf{I}(y > d^{-1})$, and we set $x = y^{-1}$. In order to sample from the last density, we add one positive latent variable m such that $[y, m] \propto \exp(-by)\mathbf{I}(m < y^{a-1})\mathbf{I}(y > d^{-1})$. Then using the Gibbs sampler we take values from the full conditional densities which are of known form:

- $y|m \equiv \exp(-by)\mathbf{I}(y > k)$ where $k = \max_x(m^{(a-1)/2}, d^{-1})$. In order to sample from this distribution see Damien *et al.* (1999).
- $m|y \sim \mathbf{U}(0, y^{a-1})$

Acknowledgments

This work has been partially supported by EU TMR network ERB-FMRX-CT96-0095 on "Computational and Statistical methods for the analysis of spatial data". Stefanos G. Giakoumatos is granted by State Scholarship Foundation of Greece. The authors want to thank an anonymous referee for very constructive comments that greatly improved the paper.

References

Besag J.E. and Green P.J. 1993. Spatial statistics and Bayesian computation (with discussion). Journal of Royal Statistical Society, Series B, 55: 25–37.

Bollerslev T., Engle R.F., and Nelson D.B. 1994. ARCH models. In: Engle R.F. and McFadden D.L. (Eds.), The Handbook of Econometrics, Amsterdam, North-Holland, Vol. 4, pp. 2959–3038.

Brooks S.P. 1999. Bayesian Analysis of Animal Abundance Data via MCMC. In: Bernardo J.M., Berger J.O., Dawid A.P., and Smith A.F.M. (Eds.), Bayesian Statistics 6. Oxford University Press, London, pp. 723–731.

Chib S. and Greenberg E. 1994. Bayes inference in regression models with ARMA(p.q) errors. Journal of Econometrics 64, pp. 183–206.

Chib S. and Greenberg E. 1995a. Understanding the metropolis-hastings algorithm. The American Statistician 49: 327–335.

Chib S. and Greenberg E. 1995b. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. Journal of Econometrics 68: 339–360.

Chib S. and Carlin B.P. 1999. On MCMC Sampling in Hierarchical longitudinal Models, Statistics and Computing 9: 17–26.

Damien P., Wakefield J., and Walker S., 1999. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. Journal of Royal Society, Series B 61: 331–344.

Edwards R.G. and Sokal A.D. 1988. Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo Algorithm. Physical Review Letters 38: 2009–2012.

Engle R.F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica 50: 987–1008.

Fiorentini G., Sentana E., and Shephard N. 2004. Likelihood-based estimation of latent generalised ARCH structures. Econometrica 72: 1481–1517.

Gelfand A.E., Smith A.F.M., and Lee T. 1992. Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. Journal of the American Statistical Association 87: 523–532.

Ghysels E., Harvey A.C., and Renault E. 1996. Stochastic volatility. In: Handbook of Statistics 14, Rao C.R. and Maddala G.S. (Eds.), Statistical Method in Finance, North-Holland, Amsterdam, pp. 119–191.

Giakoumatos S.G. 2004. Auxiliary variable samplers for time-varying volatility models. Ph.D Thesis, Department of Statistics, Athens University of Economics and Business.

Giakoumatos S.G., Vrontos I.D., Dellaportas P., and Politis D.N. 1999. An MCMC Convergence Diagnostic using Subsampling. Journal of Computational and Graphical Statistics 8(3): 431–451.

Harvey A., Ruiz E., and Sentana E., 1992. Unobserved component time series models with ARCH disturbances. Journal of Econometrics 52: 129–157.

Hastings W.K. 1970. Monte Carlo Sampling methods using markov chains and their applications. Biometrika 57: 97–109.

Higdon D.M. 1998. Auxiliary variable methods for Markov chain Monte Carlo with applications. Journal of the American Statistical Association 93: 585–595.

Jacquier E., Polson N.G., and Rossi P.E. 1994. Bayesian analysis of the stochastic volatility models (with discussion). Journal of Business and Economics Statistics 12: 371–417.

Kim S., Shephard N., and Chib S., 1998. Stochastic Volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65: 361–393.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., and Teller E. 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21: 1087–1092.

Mira A. and Tierney L. 2002. Efficiency and convergence properties of slice samplers. Scandinavian Journal of Statistics 29: 1–12.

Neal R.M. 1993. Probabilistic inference using Markov Chain Monte Carlo methods, Techinal Report CRG-TR-93-1. Department. of Computer Science, University of Toronto.

Neal R.M. 2003. Slice sampling (with discussion). Annals of Statistics 31: 705–767.

Pitt M.K. 1997. Bayesian methods for non-Gaussian state space models, D.Phil. (supervised by Neil Shephard). University of Oxford.

Pitt M.K. and Shephard N. 1999. Time-Varying covariances: A Factor Stochastic Volatility approach (with discussion). In: Bernardo J.M., Berger J.O., Dawid A.P., and Smith A.F.M. (Eds.), Bayesian Statistics 6, Oxford University Press, London, pp. 547–570.

Roberts G.O. and Rosenthal J.S. 1999. Convergence of slice sampler Markov chains, Journal of Royal Statistical Society, Series B, 61: 643–660.

Shephard N. 1996. Statistical aspects of ARCH and Stochastic Volatility. In: Cox D.R., Barndorff-Nielsen O.E., and Hinkley D.V. (Eds.), Time Series Models In Econometrics, Finance and Other Fields, Chapman and Hall, London, pp. 1–67.

Swendsen R.H. and Wang J.S. 1987. Nonuniversal critical dynamics in Monte Carlo simulations. Physical Review Letters 58: 86–88.