A BAYESIAN ANALYSIS OF LONG-MEMORY STOCHASTIC VOLATILITY

N. H. Chan
Department of Statistics, Carnegie Mellon University
Pittsburgh, PA 15213-3890, USA
e-mail: chan@stat.cmu.edu

G. Petris
Department of Mathematical Sciences, University of Arkansas
Fayetteville, AR 72701, USA
e-mail: gpetris@comp.uark.edu

October 15, 1999

Abstract

A novel Bayesian treatment of a long-memory stochastic volatility model is proposed in this paper. The model under consideration captures both stationary and nonstationary long range dependence so that it is particularly well-suited to model the empirical behavior often observed in high frequency stock market return data. By incorporating a truncated state space representation of a long-memory model, the proposed model can be expressed in a dynamic linear model framework so that posterior analyses can be conducted by Markov chain monte carlo simulation techniques. An illustration of the method is provided by the analysis of daily returns on a value-weighted index from the Center for Research in Security Prices.

1 Introduction

Let y_t be the price of an asset at time t ($t = 0, 1, \ldots$). The (continuously compounded) return on the asset at time t is defined as $r_t = \log y_t - \log y_{t-1}$. It is the average return rate over the tth time period. Stock market returns have long been observed to behave like a sequence of centered and uncorrelated random variables. This has led to model the price process as a random walk, on a log scale:

$$\log y_t = \log y_{t-1} + \eta_t,$$

where y_t denotes the market value of an asset at time t, and $\{\eta_t\}$ is a sequence of independent identically distributed random variables, typically assumed to be Gaussian. The same random walk model, taking the sampling frequency going to infinity, provides a justification for the geometric Brownian motion model used in continuous time. A closer look at return data, however, reveals that the random walk model is unrealistic. Among the reasons, the following stylized facts can be singled out:

1. High (low) returns, in absolute value, tend to cluster in periods of high (low) market volatility;
2. The squares of the returns exhibit a slowly decaying serial correlation;
3. The distribution of the returns has fatter tails than a Gaussian distribution.

In an attempt to explain these stylized facts, two basic families of models have been proposed: Generalized Autoregressive Conditionally Heteroskedastic (GARCH) models and Stochastic Volatility (SV). In the GARCH class of models, introduced by Engle (1982) and Bollerslev (1986), the conditional distribution of the return at time t, given the past returns, is assumed to be Gaussian with zero mean and variance

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2,$$

r_t being the return at time t. Estimation of the parameters is customarily done using quasi-maximum-likelihood procedures. Several extensions of the basic model have also been proposed, including the so-called E-GARCH, T-GARCH, and the semiparametric GARCH, to name a few. For an account on recent developments of GARCH models, see Bollerselv, Engle and Nelson (1994). Note that in all of these models, the volatility σ_t^2 is assumed to be a deterministic function of the past or, in probabilistic terms, a predictable process.

A stochastic volatility model, introduced by Taylor (1986), provides an alternative in which the volatility of the returns is assumed to be a non-observable, or latent, process. The volatility can be alternatively thought as the state vector, with its own evolution equation, of a non-linear state-space model. In the

most common applications, letting $v_t = \log \sigma_t^2$, the model has the form

$$\begin{cases} r_t = \sigma_t \varepsilon_t \\ \phi(B) v_t = \theta(B) \eta_t, \end{cases}$$

where $\phi(B)$ and $\theta(B)$ are polynomials in the backshift operator, subject to the standard stationarity and causality conditions, $\{\eta_t\}$ is a Gaussian white noise, and $\{\varepsilon_t\}$ is a sequence of independent standard normal random variables, the two sequences being independent. In a sense, SV provides a more flexible probabilistic model, with the evolution of the return process being driven by a two-dimensional white noise, rather than a one-dimensional one.

Several estimation procedures have been proposed to fit SV models. Among these, the method of moments have been receiving considerable attentions. Melino and Turnbull (1990) use a generalized method of moments, which is straightforward to implement, but not efficient. To circumvent the inefficiency, Gallant, Hsieh and Tauchen (1997) propose to use the efficient method of moments (EMM) to estimate a SV model. They show that not only is the EMM asymptotically efficient and possesses many of the desirable asymptotic properties such as consistency and asymptotic normality, the EMM also provides directions to calibrate the SV model in the event that the assumed model does not conform with the observed series. Andersen, Chung and Soresen (1999) conduct a Monte Carlo study on using the EMM for an SV model. Although the EMM scheme seems to be powerful and flexible, its success hinges on judicious choices of the auxiliary model which may be sensitive in small samples, see Andersen et al. (1999). Alternatively, Harvey, Ruiz and Shephard (1994) propose a quasi-maximum-likelihood approach, based on the transformation of the model into a linear state-space with non-Gaussian observation errors. A Bayesian approach is taken by Jacquier, Polson and Rossi (1994). Kim, Shephard and Chib (1998) suggest a Markov Chain Monte Carlo (MCMC) sampling based framework for Bayesian and maximum-likelihood estimator, while Sandmann and Koopman (1998) propose a maximum likelihood Monte Carlo approach. A survey on related developments in SV models can be found in Ghysels, Harvey and Renault (1995).

Fitting a GARCH or SV model to actual return data typically results in estimates of the parameters close to the boundary of the stationarity region. This is consistent with the slow decay observed in the correlation function of the squared returns. To explicitely model the persistence in the volatility, non-stationary integrated versions of each class of models have been used. More recently, stationary fractionally integrated versions have been proposed, see, for example, Robinson (1991), Harvey (1993), Ding, Granger and Engle (1993), and Baillie, Bollerslev and Mikkelsen (1996) for the GARCH case and Breidt, Crato and de Lima (1998) for the SV case.

In this paper, a new Bayesian approach is proposed to estimate a LMSV model. By means of the truncated likelihood method given in Chan and Palma (1998), a LMSV model is expressed as a linear state space formulation in terms of a dynamic linear model, introduced in West and Harrison (1997). The dynamic

linear model offers a natural platform to conduct MCMC Bayesian estimation for LMSV models. Although the proposed method is conceptually similar to the Bayesian method discussed in Jacquier et al. (1994), there remains several fundamental differences. First, our method offers a direct parametrization of a LMSV model which was not covered in Jacquier et al.. Second, the dynamic linear model set up allows one to deal with nonstationarity directly. It is not necessary to prefilter the series before fitting. This may turn out to be an important property since many of the financial data exhibit nonstationary behavior in addition to the long-memory phenomenon. Third, the truncated likelihood method offers a convenient basis to sample the posterior directly for some components of the state, although detailed Hastings-Metropolis algorithms are still required for other components. Together, our approach offers an efficient method in estimating a LMSV model.

The setup of the paper is as follows. After defining the model in section 2, we describe the prior distribution and the Markov Chain Monte Carlo simulation scheme used to sample from the posterior in section 3. The proposed methodology is then applied, in section 4, to a series of daily returns on a value-weighted index from the Center for Research in Security Prices. Finally, section 5 contains concluding remarks and suggestions for future research.

2 Long-Memory Stochastic Volatility Models

The basic setup is the stochastic volatility model:

$$\begin{cases} r_t = \sigma_t \varepsilon_t \\ (1-B)^d \phi(B)(v_t - \mu) = \theta(B)\eta_t \end{cases} \tag{1}$$

with $v_t = \log \sigma_t^2$. In (1), r_t is the return observed at time t, $\{\varepsilon_t\}$ is a sequence of independent standard Normal random variables, μ is a real constant, $d \in (-0.5, 0.5)$, $\{\eta_t\}$ is Gaussian white noise with variance τ, $\phi(\cdot)$ and $\theta(\cdot)$ are polynomials of order p, q, respectively, with all their roots outside the unit circle and with no common root, and B is the backshift operator. The model assumptions imply that v_t has a moving average representation in terms of the white noise η_t. One can truncate the infinite moving average to a finite number of terms M, say, to obtain an approximate representation of v_t. Chan and Palma (1998) showed that a better approximation can be obtained by considering the corresponding truncation of the moving average representation of the first difference of v_t:

$$\begin{aligned} &(1-B)^{d-1}\phi(B)\Delta v_t = \theta(B)\eta_t \qquad (2)\\ &\Delta v_t = (1-B)^{-d+1}\big(\phi(B)\big)^{-1}\theta(B)\eta_t \\ &\quad = \Big(\sum_{j=0}^{\infty}\varphi_j B^j\Big)\eta_t \approx \Big(\sum_{j=0}^{M}\varphi_j B^j\Big)\eta_t = \sum_{j=0}^{M}\varphi_j \eta_{t-j}. \end{aligned}$$

The coefficients φ_j of the truncated moving average representation of (Δv_t) can be computed as follows. Let $(1-z)^{-d+1} = \sum \gamma_j z^j$, i.e. $\gamma_j = (-1)^j \binom{-d+1}{j}$. (The γ_j can be conveniently computed recursively from the relation

$$\gamma_j = \frac{(d-1)d(d+1)\dots(d+j-2)}{j!} \tag{3}$$

$$= \gamma_{j-1} \frac{d+j-2}{j}, \tag{4}$$

starting with $\gamma_0 = 1$.) Then

$$\Big(\sum_{j=0}^{p} \phi_j z^j\Big)\Big(\sum_{j=0}^{\infty} \varphi_j z^j\Big) = \Big(\sum_{j=0}^{\infty} \gamma_j z^j\Big)\Big(\sum_{j=0}^{q} \theta_j z^j\Big) \tag{5}$$

$$\sum_{j=0}^{\infty} \sum_{r+s=j} \phi_r \varphi_s z^j = \sum_{j=0}^{\infty} \sum_{r+s=j} \gamma_r \theta_s z^j. \tag{6}$$

Equating the coefficient of the power series on the two sides gives the set of equations

$$\varphi_j = \gamma_j + \sum_{r=1}^{q \wedge j} \theta_r \gamma_{j-r} - \sum_{r=1}^{p \wedge j} \phi_r \varphi_{j-r}, \tag{7}$$

which can be successively solved for $j = 0, 1, \dots, M$. Note that the φ_j's are functions of d, $\underline{\phi} = (\phi_1, \dots, \phi_p)$ and $\underline{\theta} = (\theta_1, \dots, \theta_q)$ (to achieve identification, we assume throughout that $\phi_0 = \theta_0 = 1$). Let $x_t = \log r_t^2 - \log r_{t-1}^2$ and $u_t = \log \varepsilon_t^2 - \log \varepsilon_{t-1}^2$. Then (1), taking logs and differencing, implies:

$$x_t = w_t + u_t, \tag{8}$$

with $w_t = v_t - v_{t-1}$. We have therefore the following approximate model:

$$\begin{cases} x_t = w_t + u_t \\ w_t = \displaystyle\sum_{j=0}^{M} \varphi_j \eta_{t-j} \\ u_t = \log \varepsilon_t^2 - \log \varepsilon_{t-1}^2. \end{cases} \tag{9}$$

This can be conveniently represented in terms of a Dynamic Linear Model (DLM) in the sense of West and Harrison (1997), as we are going to describe next. Since w_t and u_t are independent moving average processes, each one has a state-space representation. The two can be combined to give:

$$\begin{cases} \Psi_{t+1} = \left[\begin{array}{c|cc} 0 & 0 & \\ \hline 0 & 0 & 0 \\ & \mathbf{I}_M & 0 \end{array}\right] \Psi_t + \begin{bmatrix} -\log \varepsilon_t^2 \\ \eta_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ x_t = \begin{bmatrix} 1 & \varphi_0 & \dots & \varphi_M \end{bmatrix} \Psi_t + \log \varepsilon_t^2 \end{cases} \tag{10}$$

where $\Psi_t = (\Psi_{t,1}, \ldots, \Psi_{t,M+2})$ is an $(M+2)$ dimensional unobservable (latent) state vector at time t. The DLM is completely specified once a distribution for the state vector at time $t = 0$ is given. If we imagine that the dynamic of the system can be extended into the past, the components of Ψ_1 have the following interpretation:

$$\begin{aligned} &\Psi_{0,1} = -\log \varepsilon_{-1}^2, \\ &X_{0,j} = \eta_{2-j}, \qquad j = 2, \ldots, M+2. \end{aligned} \tag{11}$$

Guided by this interpretation, we assign Ψ_0 a Normal distribution with mean $(1.27, 0, \ldots, 0)'$ and variance $\mathrm{diag}(\pi^2/2, \tau, \ldots, \tau)$. Note that -1.27 and $\pi^2/2$ are the mean and variance of the $\log \chi_1^2$ distribution. It would be very convenient to work with a Gaussian DLM. In order to do so, one can approximate the distribution of $\log \varepsilon_t^2$, which is a $\log \chi_1^2$ distribution, to a convenient accuracy with a finite mixture of normal distributions, a device introduced by Shephard (1994) (see also Kim et al.; 1998, p.371). Denoting by $\mathcal{L}(X)$ the distribution of X, for any random element X, we can write

$$\mathcal{L}(\log \varepsilon_t^2) \approx \sum_{j=1}^{N} \pi_j \mathcal{N}(m_j, \sigma_j^2), \tag{12}$$

where $\mathcal{N}(m, \sigma^2)$ denotes a Gaussian distribution with mean m and variance σ^2, and the π_j's are positive weights adding to one. This representation suggests that we add to the model a vector of T discrete independent latent variables

$$\underline{K} = \big(K_1, \ldots, K_T\big), \tag{13}$$

with distribution defined by

$$P(K_t = j) = \pi_j, \qquad j = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{14}$$

The link between this latent vector and the model is set by putting

$$P(\log \varepsilon_t^2 \le x | \underline{K}) = \frac{1}{\sigma_j \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{(\xi - m_j)^2}{2\sigma_j^2} \right) d\xi \qquad \text{if } K_t = j, \tag{15}$$

$j = 1, \ldots, N$, $t = 1, \ldots, T$. Up to the approximation (12), the marginal distribution of the sequence (u_t) has not changed; on the other hand, conditionally on $\underline{K}$, the DLM (10) is Gaussian.

3 Prior Distributions and Posterior Analysis

In roder to proceed with the posterior analysis, a prior distribution for the parameters τ, d, $\underline{\phi}$, $\underline{\theta}$ is assigned. The prior distribution for d, $\underline{\phi}$, $\underline{\theta}$ is taken to

be uniform in its domain of definition. The prior distribution of τ is taken to be inverse gamma with parameters α_τ, β_τ, with density

$$p_\tau(\xi) = \frac{{\beta_\tau}^{\alpha_\tau}}{\Gamma(\alpha_\tau)} \xi^{-\alpha_\tau - 1} \exp\left(-\frac{\beta_\tau}{\xi}\right), \qquad \xi > 0. \tag{16}$$

Furthermore, τ is assumed to be independent of d, $\underline{\phi}$ and $\underline{\theta}$.

To analyze the posterior distribution, we need to generate a sample from the joint distribution of all the parameters and the latent variables Ψ_t, condition on the observed sequence x_t, $(t = 1, \ldots, T)$. Although parametrization (10) provides a convenient DLM expression for the model, it is somehow redundant as one component of the parameter is the sequence $(\Psi_t)_{t=1,\ldots,T}$ of the state vectors of the state-space model (10). Note that each Ψ_t is an $(M+2)$-dimensional vector, the whole sequence is a point in $\Re^{T(M+2)}$. On the other hand, the innovations in (10) are only 2-dimensional and consequently $(\Psi_t : t = 1, \ldots, T)$ effectively only belongs to a proper submanifold of $\Re^{T(M+2)}$. Moreover, the observation equation in (10) imposes additional constraints on the support of the posterior distribution of the parameters. Therefore, we can formally reparametrize the model as follows.

$$\Theta = \left(\Psi_0, \Psi_{1,2}, \Psi_{2,2}, \ldots, \Psi_{T,2}\right), \tag{17a}$$

$$\Lambda = \left(\Theta, \underline{K}, \tau, d, \underline{\phi}, \underline{\theta}\right). \tag{17b}$$

Then, given the data, there is a one-to-one correspondence between the two following vectors:

$$\left((\Psi_t, t = 1, \cdots, n), \underline{K}, \tau, d, \underline{\phi}, \underline{\theta}\right) \longleftrightarrow \Lambda \tag{18}$$

One way of the bijection (the "$\rightarrow$" part) trivially consists in dropping some of the components of the vector on the left-hand side. Let us spend just a few words to describe the "$\leftarrow$" part. We must show that there is a unique way to recover $\Psi_{t,j}$, $t = 2, \ldots, T$, $j \neq 2$, from Λ and $(x_t)_{t=1,\ldots,T}$. From (10),

$$\Psi_{t,j} = \Psi_{t-1,j-1}, \qquad j = 3, \ldots, M+2, \tag{19}$$

which, knowing Ψ_0, can be solved recursively for $j = 3, \ldots, M+2$ and $t = 1, \ldots, T$. Note, incidentally, that the data are not needed to recover the $\Psi_{t,j}$'s for $j \geq 2$. As far as $\Psi_{t,1}$, is concerned, from (10) one has:

$$\Psi_{t,1} = \Psi_{t-1,1} + w_{t-1} - x_{t-1}, \qquad t = 1, \ldots, T, \tag{20}$$

with

$$w_t = \sum_{j=0}^{M} \varphi_j \Psi_{t,j+2}. \tag{21}$$

In view of the one-to-one correspondence described above, we will focus on how to obtain a sample from the conditional distribution of Λ given the data. We

will make use of a Markov Chain sampling scheme, namely a Gibbs sampler with Metropolis steps embedded. Loosely speaking, each of the six components of Λ (which may itself be multidimensional, such as Θ, for example) is sampled from its full conditional distribution, i.e., its conditional distribution given the data and the other parameters. More formally, starting from an arbitrary value $\lambda^{(0)}$ in the parameter space, at the nth iteration the sampler cycles through a set of six transition kernels:

$$Q_j(d\xi; \lambda_1^{(n)}, \ldots \lambda_j^{(n)}, \lambda_{j-1}^{(n-1)}, \ldots, \lambda_6^{(n-1)}) \qquad j = 1, \ldots, 6. \tag{22}$$

Here $Q_j(d\xi; \lambda)$ is a transition kernel which puts all the mass on the set $\{\xi : \xi_k = \lambda_k, k \neq j\}$ and for which $\mathcal{L}(\Lambda \mid (\Delta x_t))$ is an invariant distribution. For a thorough overview of Markov Chain Monte Carlo methods, we refer the interested reader to the lucid paper by Tierney (1994). We now describe now each Q_j in detail.

3.1 Sampling Θ.

To sample Θ, we generate the full vector (Ψ_t) from its conditional distribution given everything else, and retain only the coordinates of interest. Since, given all the other parameters and latent variables, the model reduces to the DLM (10), sampling from the full conditional distribution of (Ψ_t) is equivalent to sampling from the posterior distribution of the state vectors at time $t = 1, \ldots, T$ in a completely specified Gaussian DLM. This can be done efficiently using the forward filtering, backward sampling approach of Frühwirth-Schnatter (1994) (see also Shephard 1994; Carter and Kohn 1994). Once the entire sequence (Ψ_t) has been generated, we keep only $\Theta = (\Psi_0, \Psi_{1,2}, \ldots, \Psi_{T,2})$.

3.2 Sampling $\underline{K}$.

From Bayes theorem, the full conditional distribution of $\underline{K}$, or of any other of the remaining components of Λ, has a density (with respect to the appropriate dominating measure) proportional to the density of the joint distribution of Λ and the data $(x_t : t = 1, \ldots, T)$. The latter can be determined writing

$$\begin{aligned} p(\Lambda, (x_t)) &= p((x_t) \mid \Lambda)\, p(\Lambda) \\ &= p((x_t) \mid \Lambda)\, p(\Theta \mid \Lambda_{\setminus\Theta})\, p(\Lambda_{\setminus\Theta}) \end{aligned} \tag{23}$$

where $\Lambda_{\setminus\Theta}$ denotes the vector Λ without its first component Θ. The third factor in the last row of (23) is just the product of the marginal prior densities involved:

$$p(\Lambda_{\setminus\Theta}) = p(\underline{K})\, p(\tau)\, p(d)\, p(\underline{\phi})\, p(\underline{\theta}) \tag{24}$$

We will now give an explicit expression for the remaining two factors. First, note that

$$\begin{aligned} p((x_t) \mid \Lambda) &= p((\nu_t) \mid \Lambda) \\ &= \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_{K_t}^2}} \exp\left(-\frac{(\nu_t - m_{K_t})^2}{2\sigma_{K_t}^2}\right), \end{aligned} \tag{25}$$

with $\nu_t = \log \varepsilon_t^2$. The u_t's can be computed recursively from the data and Λ as follows:

$$\begin{aligned} &\nu_1 = x_1 - \Psi_{1,1} - w_1, \\ &\nu_t = \nu_{t-1} + x_t - w_t, \qquad t = 2, \ldots, T, \end{aligned} \tag{26}$$

where w_t is given by (21).

As far as $p(\Theta \mid \Lambda_{\setminus\Theta})$ goes, one has, from (10):

$$\begin{aligned} p(\Theta \mid \Lambda_{\setminus\Theta}) &= p(\Psi_1 \mid \Lambda_{\setminus\Theta}) \prod_{t=2}^{T} p(\Psi_{t,2} \mid \Lambda_{\setminus\Theta}) \\ &= \frac{1}{\pi^{3/2}} \exp\left(-\frac{(\Psi_{1,1} - 1.27)^2}{\pi^2}\right) \prod_{j=2}^{M+2} \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\Psi_{1,j}^2}{2\tau}\right) \\ &\quad \times \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\tau}} \exp\left(-\frac{\Psi_{t,2}^2}{2\tau}\right) \\ &= \textit{constant} \times \exp\left(-\frac{(\Psi_{1,1} - 1.27)^2}{\pi^2}\right) \tau^{-\frac{M+T}{2}} \\ &\quad \times \exp\left(-\frac{1}{2\tau}\left(\sum_{j=2}^{M+2} \Psi_{1,j}^2 + \sum_{t=2}^{T} \Psi_{t,2}^2\right)\right). \end{aligned} \tag{27}$$

Going back to the full conditional distribution of $\underline{K}$, considering only the factors in $p(\Lambda, (x_t))$ which depend on $\underline{K}$, one has

$$p(\underline{K}) \propto p(\Lambda, (x_t)) \propto \prod_{t=1}^{T} \frac{\pi_{K_t}}{\sigma_{K_t}} \exp\left(-\frac{(\nu_t - m_{K_t})^2}{2\sigma_{K_t}^2}\right). \tag{28}$$

Therefore the components of $\underline{K}$ can be independently drawn from the densities

$$p(K_t) \propto \frac{\pi_{K_t}}{\sigma_{K_t}} \exp\left(-\frac{(\nu_t - m_{K_t})^2}{2\sigma_{K_t}^2}\right), \tag{29}$$

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

3.3 Sampling τ.

The full conditional distribution of τ has density:

$$\begin{aligned} p(\tau \mid \dots) &\propto p(\Lambda, (x_t)) \\ &\propto \tau^{-\alpha_\tau - \frac{M+T}{2} - 1} \exp\left(-\frac{1}{\tau} \left(\beta_\tau + \frac{1}{2} \sum_{j=2}^{M+2} \Psi_{1,j}^2 + \frac{1}{2} \sum_{t=2}^{T} \Psi_{t,2}^2 \right) \right). \end{aligned} \tag{30}$$

It is therefore an inverse gamma distribution with parameters

$$\begin{aligned} \alpha'_\tau &= \alpha_\tau + \frac{M+T}{2}, \\ \beta'_\tau &= \beta_\tau + \frac{1}{2} \sum_{j=2}^{M+2} \Psi_{1,j}^2 + \frac{1}{2} \sum_{t=2}^{T} \Psi_{t,2}^2. \end{aligned} \tag{31}$$

3.4 Sampling d, $\underline{\phi}$ and $\underline{\theta}$.

The full conditional densities of the remaining parameters do not have an analytic form which can be recognized as corresponding to any known and well-studied distribution. Therefore to draw from these distributions we use Metropolis-Hastings algorithm (see Tierney (1994) for background and details). For all the parameters in this group, the density to be drawn is proportional to

$$\begin{aligned} p(\Lambda, (x_t)) &\propto p(d)\, p(\underline{\phi})\, p(\underline{\theta}) \prod_{t=1}^{T} \frac{1}{\sigma_{K_t}} \exp\left(-\frac{(\nu_t - m_{K_t})^2}{2\sigma_{K_t}^2} \right) \\ &= \prod_{t=1}^{T} \frac{1}{\sigma_{K_t}} \exp\left(-\frac{(\nu_t - m_{K_t})^2}{2\sigma_{K_t}^2} \right). \end{aligned} \tag{32}$$

Note that the parameters appear in the preceding expression through the ν_t's. For each one-dimensional full conditional, the proposal distribution we use is based on a linear approximation of the logarithm of the target density (32), as described in Gilks and Wild (1992)

4 An Application

We apply the model and estimation technique described in the previous section to a financial time series. The data consists of the daily returns for the value-weighted market index from the Center for Research in Security Prices from July 1962 to July 1989. Following common practices, correlations in the return data due to the day of the week and month of the year were removed using standard filters, details can be found in Breidt et al. (1998). The series of these returns, together with the series of log squared returns, are plotted in Figure 1. There seems to be an increasing trend in the series, suggesting strong persistence or

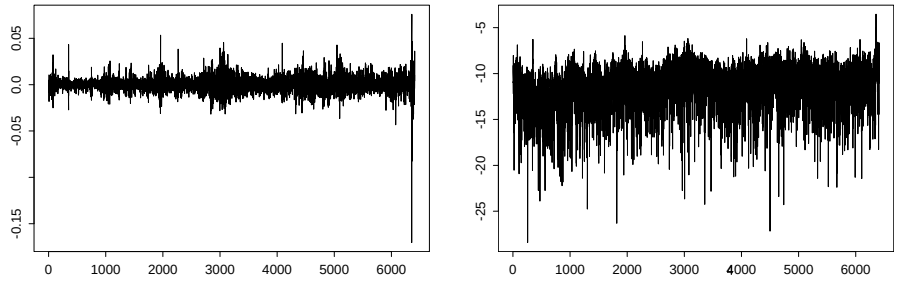

Figure 1: Daily returns (left) and log squared returns (right).

even nonstationarity. Fitting a straight line to the data by ordinary least squares gives a t-value of 15 for the slope parameter. Since the standard normality assumptions clearly do not hold in the present situation, it is difficult to interpret this number, for example attaching a p-value to it. However, we are inclined to judge it as a "large number", though informally, prompting the use of a model that allows a nonstationary behavior.

We model the return data as

$$\begin{cases} r_t = \sigma_t \varepsilon_t, \\ (1-B)^d (1-\phi_1 B)(v_t - \mu) = \eta_t, \end{cases} \tag{33}$$

using the prior described in the previous section. Posterior summaries (the mean and four quantiles) for selected parameters resulting from the MCMC simulation are reported in Table I.

	0.05	0.25	Mean	0.75	0.95
d	0.555	0.642	0.675	0.717	0.722
ϕ_1	0.589	0.590	0.595	0.596	0.602
τ	0.000129	0.000229	0.002655	0.005216	0.077035

Table I.

The posterior distribution of d reflects the earlier observation that there exists nonstationarity in the volatility process.

One advantage of the Bayesian approach is that a full posterior distribution is available, so that inference on events or quantities depending on the parameters is conceptually straightforward. For example, one issue with this kind of data is whether the process driving the volatility is stationary or not. This formally corresponds to test the hypothesis that d is less than 0.5. For this data set, the posterior probability that the process is nonstationary ($d \geq 0.5$), evaluated from the Monte Carlo sample, turns out to be 99%. Note that the prior we use is noninformative with respect to this issue, in the sense that $P(-0.5 < d < 0.5) = P(0.5 \leq d < 1.5) = 1/2$.

Breidt et al. (1998), using an approach based on the spectral likelihood, find estimates $\hat{\phi}_1 = 0.932$ and $\hat{d} = 0.444$. It should be pointed out, in order to explain the discrepancy between those estimates and the corresponding posterior means reported in Table 1, that we are using a different model. In fact, even if the set of equations describing the observation process and the evolution of the volatility is the same, and is expressed in terms of the same parameters, our parameter space is different. While we allow the long memory parameter d to vary in $(-0.5, 1.5)$, Breidt et al. (1998) constrain this parameter to the stationarity region $(-0.5, 0.5)$.

5 Concluding Remarks

Returns on stocks or indexes typically show a nonlinear behavior. Several models have been proposed to describe this kind of data, usually assuming the unobservable volatility of the returns to follow either a stationary process (e.g., GARCH, SV), or a nonstationary one (e.g., IGARCH). The present paper, after reviewing the most popular models and estimation techniques for financial return data, introduces a model that encompasses stationarity and nonstationarity, as well as long-range dependence, a feature frequently observed in daily financial time series. The Bayesian approach taken here allows one, by combining a (typically noninformative) prior with the evidence provided by the data through the likelihood function, to obtain a readily interpretable posterior probability of the volatility process being stationary. In the example considered in Section 4, the evidence against stationarity is fairly strong. This is in accord with the recent findings that when daily returns are analyzed, one often ends up with an IGARCH(1,1) model or an SV model with parameters close to the boundary of the stationarity region.

A stylized fact about daily returns that has not been considered here is the excess kurtosis in the returns. This can be easily accommodated in our model by taking the distribution of ξ_t in equation (1) to be a Student's t with fixed degrees of freedom. Then in the mixture of normals (12), the weights π_j's, means m_j's, and variances σ_j^2's have to be revised so that the mixture approximate the corresponding moments of the $\log(t^2)$ distribution. Note that this does not make the analysis or the simulation scheme more involved. In a more general framework, one should be able to estimate the extent to which the ξ_t's are leptokurtic. One possibility is to consider for ξ_t a Student's t distribution with unknown degrees of freedom ν for a finite number of possible values of ν. This would only add one extra discrete distribution to sample in the MCMC step.

Several other topics remain open for future research, including the important problem of forecasting the volatility for risk management. The state-space formulation, together with the simulation approach, is perfectly suited to generate future paths of the volatility from the appropriate predictive distribution. Generating a stretch of future volatilities for a fixed value of the parameters ϕ_j in equation (2), determined by the current state of the chain, is as easy as generating from a moving average process with known parameters. From

these future volatility scenarios, one can compute means, probability intervals, standard deviations. The added value brought in by the simulation approach is that one can also look at typical future behaviors of the volatility, in addition to pointwise summaries such as means and histograms. Clearly, one needs to check the predictive properties of the model against past data before using these predictions.

References

Andersen, T., Chung, H. and Soresen (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study, *Journal of Econometrics* **91**: 61–87.

Baillie, R., Bollerslev, T. and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **74**: 3–30.

Bollerselv, T., Engle, R. and Nelson, D. (1994). ARCH models, *in* R. Engle and D. McFadden (eds), *Handbook of Econometrics, Volume IV*, North-Holland.

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

Breidt, F., Crato, N. and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* **83**: 325–348.

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

Chan, N. and Palma, W. (1998). State space modeling of long-memory processes, *The Annals of Statistics* **26**: 719–740.

Ding, Z., Granger, C. and Engle, R. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance* **1**: 83–106.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation, *Econometrica* **50**: 987–1008.

Frühwirth-Schnatter, S. (1994). Data augmentation and Dynamic Linear Models, *Journal of Time Series Analysis* **15**: 183–202.

Gallant, A., Hsieh, D. and Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* **81**: 159–192.

Ghysels, E., Harvey, A. and Renault, E. (1995). Stochastic volatility, *in* G. Maddala (ed.), *Statistical Methods in Finance*, Vol. 14 of *Handbook of Statistics*, North-Holland.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics* **41**: 337–348.

Harvey, A. (1993). Long memory in stochastic volatility, *Technical report*, London School of Economics.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**: 247–264.

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

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

Melino, A. and Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* **45**: 239–265.

Robinson, P. (1991). Testing for strong serial correlation and dynamics conditional heteroskedasticity in multiple regression, *Journal of Econometrics* **47**: 67–84.

Sandmann, G. and Koopman, S. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* **87**: 271–301.

Shephard, N. (1994). Partial non-Gaussian state-space, *Biometrika* **81**: 115–131.

Taylor, S. (1986). *Modeling financial time series*, Wiley.

Tierney, L. (1994). Markov chains for exploring posterior distributions, *The Annals of Statistics* **22**: 1701–1762.

West, M. and Harrison, J. (1997). *Bayesian forecasting and dynamic models*, 2nd edn, Springer.