SEMIPARAMETRIC BAYESIAN INFERENCE OF LONG-MEMORY STOCHASTIC VOLATILITY MODELS

By Mark J. Jensen

Brigham Young University

First Version received January 2003

Abstract. In this paper, a semiparametric, Bayesian estimator of the long-memory stochastic volatility model's fractional order of integration is presented. This new estimator relies on a highly efficient, Markov chain Monte Carlo (MCMC) sampler of the model's posterior distribution. The MCMC algorithm is set forth in the time-scale domain of the stochastic volatility model's wavelet representation. The key to and centerpiece of this new algorithm is the quick and efficient multi-state sampler of the latent volatility's wavelet coefficients. A multi-state sampler of the latent wavelet coefficients is only possible because of the near-independent multivariate distribution of the long-memory process's wavelet coefficients. Using simulated and empirical stock return data, we find that our algorithm produces uncorrelated draws of the posterior distribution and point estimates that rival existing long-memory stochastic volatility estimators.

Keywords. Dirichlet process prior; long memory; Markov chain Monte Carlo; Metropolis–Hastings; semiparametric; stochastic volatility; wavelets.
JEL classification numbers. C11, C14, C22.

1. INTRODUCTION

A growing area of the time-varying heteroskedasticity literature is stochastic volatility (Taylor, 1994; Ghysels *et al.*, 1996; Shephard, 1996), in particular, stochastic volatility models whose autocovariance function of which displays the slow geometric decay commonly found in the correlogram of a financial asset's level of risk (Andersen and Bollerslev, 1997). This strong persistence in volatility has recently been modelled with the fractionally integrated class of long-memory processes (Adenstedt, 1974; McLeod and Hipel, 1978; Granger and Joyeux, 1980; Hosking, 1981). Proposed independently by Breidt *et al.* (1998) and Harvey (1998), the long-memory stochastic volatility (LMSV) model accurately reflects the slow decay found in the empirical autocorrelation of volatility.

Estimation of the LMSV model, however, has been a very complicated endeavour. In addition to the complexities associated with the high-dimensional integral calculation of the stochastic volatility's likelihood function, the persistent nature of the fractionally integrated model requires projecting the LMSV model into the frequency domain so that the quasi-likehood function can be computationally feasible. Like its predecessor, the first-order autoregressive stochastic volatility

0143-9782/04/06 895–922 JOURNAL OF TIME SERIES ANALYSIS Vol. 25, No. 6

model (Taylor, 1986, p. 74; Scott, 1987), the LMSV model has been estimated with method-of-moment-type estimators (Wright, 1999) and quasi-maximum likehood estimators (Breidt *et al.* 1998). The LMSV model's fractional differencing parameter has also been estimated with the Geweke and Porter-Hudak's (1983) (GPH) semiparametric, log-periodogram estimator (Deo and Hurvich, 2001).

A Bayesian estimator designed around an algorithm that samples from the intractable posterior distribution of the LMSV parameter vector augmented with the unobservable volatilities eliminates the need to compute the likelihood function. Inference with a Bayesian estimator also does not depend on the estimator's asymptotic properties. Instead, the sampling algorithm produces a Monte Carlo draw of the LMSV posterior distribution from which exact small-sample inference can be made. To our knowledge, such a Bayesian sampler of the LMSV model within the wavelet domain does not currently exist.[1]

The main contribution of this paper is the construction of such a sampling algorithm. We accomplish this by designing a fast and efficient Markov chain Monte Carlo (MCMC) simulator in the wavelet domain that converges quickly to the target density and produces a mix of draws from the desired posterior distribution. Our MCMC algorithm augments the latent volatility's wavelet coefficients with the LMSV parameters. Realizations from the unknown joint posterior distribution are then produced by iteratively drawing the parameters and latent volatility's wavelet coefficients from their conditional distributions where the conditioning is on the observable returns wavelet coefficients and the current value of the other unknowns.

This paper's sampling algorithm is innovative in that the long-memory process' nearly independent wavelet coefficients are used to design a sampler of the latent wavelet coefficients' multivariate conditional distribution; i.e. we sample the entire vector of latent wavelet coefficients conditional on the data and the value of the parameters in a single draw. Because a MCMC algorithm draws the latent volatilities conditionally on the value of the other unknowns, in the time domain the highly persistent behaviour of long-memory volatility causes the MCMC draws to be strongly correlated when each individual observation of volatility is drawn conditionally on the value of the other volatilities and unknowns. Such strong degrees of correlation between draws is undesirable since it leads to an inefficient sampler that fails to mix well. In other words, the algorithm becomes stuck in a particular portion of the posterior distribution and is unable to sample from across the entire distribution.

Our wavelet domain sampler of the LMSV model does not suffer from this inefficiency. Wornell and Oppenheim (1992), Tewfik and Kim (1992), and Dijkerman and Mazumdar (1994) proved for continuous time-fractional Brownian motion processes and McCoy and Walden (1996), and Jensen (1999a,b, 2000) for discrete fractionally integrated series, that a long-memory process' wavelet coefficients are 'nearly' uncorrelated. This approximately independent behaviour by the latent volatility's wavelet coefficients enables our sampler to be quick, efficient and capable of producing uncorrelated draws from the targeted posterior distribution.

We apply our sampler to artificially generated LMSV data and daily stock return data. For the artificially generated LMSV data our Bayesian estimator produces smaller bias and mean-squared error (MSE) of the differencing parameter than the GPH estimator, and a smaller MSE than the quasi-maximum likelihood estimator. Empirically, we find strong evidence of long-memory volatility when the Bayesian estimator is applied to over 30 years of daily stock returns for the S&P 100 company Aluminum Corporation of America (AA).

The contents of the paper are as follows. In Section 2 we introduce the long-memory stochastic volatility model along with its unique property of long-term persistence as quantified by the latent volatility process' autocovariance and power spectrum. Section 3 constructs the MCMC simulator for a mixture model of the log-squared return's wavelet coefficients. In Section 3 we also provide the necessary wavelet theory to construct our sampling algorithm. We would recommend either Mallat (1999) or Percival and Walden (2000) to those readers interested in obtaining a more complete understanding of wavelet analysis. Examples of the semiparametric Bayesian wavelet estimator as applied to simulated LMSV models and daily compounded stock return data are found in Section 4. Our conclusions are presented in Section 5.

2. MODEL

Define the long-memory stochastic volatility model as:

$$\begin{aligned} y_t &= \sigma \exp\{h_t/2\}\xi_t \\ (1-L)^d h_t &= \sigma_\eta \eta_t, \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where at time t the mean corrected return from holding a financial instrument is y_t, and h_t is the unobservable log volatility that behaves as a fractionally integrated process.[2] We assume $|d| < 1/2$, and ignore the possible presence of leverage effects (see Nelson, 1991; Harvey and Shephard, 1996) by assuming that the innovations ξ_t and η_t are uncorrelated standard normal white noise processes.[3] The fractional differencing operator is $(1 - L)^d$, where L is the lag operator, and $x_{t-s} = L^s x_t$, is defined by its binomial expansion. The parameter σ_η is the standard deviation of the log-volatility and σ is the modal instantaneous volatility. Finally, we denote the parameters of the LMSV model by the three-dimensional vector $\theta = (\sigma, \sigma_\eta, d)'$.

The fractionally integrated process, h_t, was independently introduced by Granger and Joyeux (1980) and Hosking (1981).[4] Since its inception, the fractionally integrated process has been associated with long-memory behaviour as quantified by the slow hyperbolic rate of decay in h_t's autocovariance function:

$$E[h_t h_{t+s}] \sim |s|^{2d-1} \quad \text{as } s \to \infty,$$

where $a_s \sim b_s$ means $a_s/b_s \to c$, for some positive constant c, as $s \to \infty$, and the pole at frequency zero in its power spectrum:

$$S(\omega) \sim |\omega|^{-2d} \quad \text{as } \omega \to 0.$$

These quantitative properties of h_t are equivalent to just two of several possible definitions of long memory (Beran 1994, p. 42). When $-1/2 < d < 0$, h_t is said to be anti-persistent or an intermediate-memory process, since the autocorrelation function will be negative for all non-zero arguments and will decay hyperbolically to zero. Whereas, when $0 < d < 1/2$, h_t exhibits the strong correlation and highly persistent behaviour of long memory as quantified by the divergence, $\sum_s |E[h_t h_{t+s}]| \to \infty$. Because slow hyperbolic decay in the autocorrelation and strong persistence are the prevalent features of financial volatility data and is solely dependent on d being positive, our interest in this paper is in providing a semiparametric, Bayesian estimator of d when the true value of d is found in the interval (0,1/2).

3. BAYESIAN ESTIMATION OF THE LMSV

The estimation problem one encounters with the entire class of stochastic volatility models, regardless of the latent volatility following a short- or long-memory process, is the general intractable nature of a nonlinear latent model's likelihood function. Because every observation of y_t is comprised of the two innovations, ξ_t and η_t, the likelihood function is a mixture over the distribution of the latent h:

$$f(y|\theta) = \int f(y|h,\theta) f(h|\theta)\, \mathrm{d}h,$$

where $y = (y_1,\ldots,y_T)'$ and $h = (h_1,\ldots,h_T)'$.

As the above likelihood function does not have a solution, both the maximum likelihood estimator of θ and the Bayesian posterior distribution, $\pi(\theta|y) \propto \pi(\theta) f(y|\theta)$, where $\pi(\theta)$ is the parameter vector's prior, are untenable. In the Bayesian context, this problem is overcome by augmenting θ with the unobservable volatilities, h, and sampling from $\pi(\theta,h|y)$ (see Tanner and Wong, 1987; Tierney, 1994). Given M realizations, $\theta^{(i)}$ and $h^{(i)}$, $i = 1,\ldots,M$, from $\pi(\theta,h|y)$, it follows that inference about $\pi(\theta|y)$ can be made from these draws since:

$$\pi(\theta|y) = \int f(\theta|h,y) f(h|y)\, \mathrm{d}h \approx M^{-1} \sum_{i=1}^{M} f(\theta|h^{(i)},y) f(h^{(i)}|y),$$

and by the law of iterative expectation (see Gelfand and Smith, 1990):

$$M^{-1} \sum_{i=1}^{M} \theta^{(i)} \to E[\theta|y], \quad \text{as } M \to \infty.$$

Equation (1) can be transformed into an infinite dimensional state- space model (see Chan and Palma 1998) by squaring the returns, y_t, and taking logs, to obtain the stochastic volatility model's linear offset representation:

$$y_t^* = \log \sigma^2 + h_t + z_t, \tag{2}$$

where $y_t^* = \log(y_t^2 + c)$, $z_t = \log \xi_t^2 + 1.2704$, and $\log \xi_t^2$ is distributed $\log \chi_{(1)}^2$ with mean -1.2704 and variance $\pi^2/2$. The offset constant c is introduced to reduce the explosive nature of the log-squared returns that occurs when returns are close to or equal to zero (see Fuller 1996, pp. 495–496; Kim *et al.* 1998). Throughout this article we will stick to the offset representation with $c = 0.0005$.

Because of the intractable nature of the stochastic volatility model's likelihood function, a Bayesian estimator of the LMSV model's d amounts to designing a MCMC simulator that makes draws of θ and h from the augmented posterior distribution $\pi(\theta,h|y^*)$. (For a general introduction to MCMC simulation see Chib, 2001; Chib and Greenberg, 1996.) To understand the workings of a MCMC algorithm for the LMSV model better, we define the blocks of unknowns to be θ and h. Each block's conditional distribution is dependent on y^* and the current value of the other block; i.e. draws are made from $\pi(\theta|y^*,h)$ and $\pi(h|y^*,\theta)$. By Bayes theorem, the form of these conditional distributions depend on the distribution of y^*. However, since z_t in Equation (2) is i.i.d. $\log \chi_{(1)}^2$, both $\pi(\theta|y^*,h)$ and $\pi(h|y^*,\theta)$ are intractable.

In the context of a first-order autoregressive (AR) stochastic volatility model, Kim *et al.* (1998) overcame the intractable nature of the conditional distributions by approximating the $\log \chi_{(1)}^2$ distribution with a mixture of independent Gaussian distributions:

$$z_t|s_t = i \sim \mathcal{N}(\mu_i - 1.2704, \sigma_i^2), \quad i = 1, \ldots, K, \tag{3}$$

where the number of mixture components, K, probability masses, π_i, means, μ_i, and variances, σ_i^2, are set equal to values that cause the first four moments of the mixture distribution to match those of the $\log \chi_{(1)}^2$ distribution. By approximating the $\log \chi_{(1)}^2$ innovations with a mixture of independent normals, Kim *et al.* (1998) is able to draw $\theta|y^*,h$ and $h|y^*,\theta$ from tractable distributions. The cost of the mixture approximation is the unknown states, s_t, which Kim *et al.* (1998) augment with the other unknowns and defines as a third block, $s|y^*,h,\theta$, in their MCMC simulator. We will use the mixture approach of Kim *et al.* (1998) in our sampler of the LMSV and hence, augment our unknowns with the mixture state vector, s. However, we do not fix the mixture parameters, instead, we allow the data to determine them a manner explained in Section 3.2.

Sampling from $\theta|y^*,h,s$, and $s|y^*,\theta,h$ is fairly straightforward. Sampling $h|y^*,\theta,s$, however, is a challenge, especially since h_t is a long-memory process. If h_t were a first-order autoregressive process, draws from $h|y^*,\theta,s$ could be performed with the de Jong and Shephard (1995) multi-state simulation smoother. The simulation smoother sampler of $h|y^*,\theta,s$ is a backward recursion algorithm based on the Kalman filter that has been successfully

applied to short-memory univariate (Kim *et al.*, 1998) and multivariate (Chib *et al.*, 2002a; 2002b) stochastic volatility models.

Unfortunately, the simulation smoother is not viable when h_t follows a long-memory process. The reasons is that as a Kalman filter-based sampler, the simulation smoother of a long-memory process requires the calculation and storage of the $T \times T$ gain matrix (Chan and Palma, 1998, Thm 2.2). Both the calculation and storage of the gain matrix amounts to a $\mathcal{O}(T^3)$ algorithm; one in terms of computing cycles, the other in computer memory. Given the large number of observations commonly found in financial data, the simulation smoother of a long-memory process will be so computationally taxing and memory intensive, that when it is integrated into the MCMC sampler, the overall computing cost will be prohibitive.

Faced with the computing and memory costs of the long-memory simulation smoother, we choose to utilize the favourable properties of the long-memory process' wavelet coefficients to design a new, fast, and efficient multi-state sampler of the latent long-memory volatility. We use the orthonormal wavelet representation of the latent volatility process for the same reasons. Jensen (1999b, 2000) found the wavelet to be useful in analysing long-memory processes; the wavelet coefficients of a fractionally integrated process is a near-independent, multivariate, Gaussian process the variance of which is of a known functional form.

3.1. *Wavelet representation of LMSV*

Wavelets are a redundant family of self-similar basis functions in the sense that all the basis functions have the same shape but are stretched, compressed, and time-shifted versions of one another. This property of the wavelet results in the fractionally integrated process's wavelet coefficients being a nearly independent self-similar process the variance of which is equal to a power function of d.

Before constructing our multi-state sampler of the latent wavelet coefficients, we first provide the necessary theory behind the wavelet that is needed to understand how the sampling algorithm works.[5] Let $\psi(t)$ be a real-valued function that is a member of the set of square integrable functions $\mathscr{L}^2(\mathbb{R})$, is centered near 0, has a norm equal to 1, and satisfies the wavelet admissibility condition, $\int \psi(t) \quad \mathrm{d}t = 0$; i.e. ψ oscillates and rapidly converges to 0 as $t \to \pm\infty$. In the Fourier domain, an equivalent admissibility condition is $2\pi \int |\hat{\psi}(\omega)|^2 |\omega|^{-1} \mathrm{d}\omega < \infty$, where $\hat{\psi}$ is the Fourier transform of ψ. Since $\psi \in \mathscr{L}^2$, the function $\hat{\psi}$ satisfies the inequality $|\hat{\psi}(\omega)| \leq C(1 + |\omega|)^{-2}$, for some $C < \infty$. So the frequency domains' admissibility condition will be satisfied if $\hat{\psi}(0) = \int \psi(t)\, \mathrm{d}t = 0$.[6]

The function ψ is called the mother wavelet and from the admissibility condition ψ can be viewed as a well-localized function in both time and frequency space; i.e. in time, $\psi(t) \to 0$ as $t \to \pm\infty$, and in frequency, $|\hat{\psi}(\omega)| \leq C(1 + |\omega|)^{-2}$, for $C < \infty$. In the time domain, ψ is centered near 0 and in the frequency domain, the energy of $\hat{\psi}$ is effectively concentrated on

$[-2\pi,-\pi) \cup (\pi,2\pi]$. An example of a wavelet is the Shannon wavelet (Hernández and Weiss, 1996, p. 61):

$$\psi(t) = -2\frac{\sin(2\pi t) + \cos(\pi t)}{\pi(2t+1)},$$

the Fourier transform of which is:

$$\hat{\psi}(\omega) = \begin{cases} e^{i\omega/2}, & \text{if } \omega \in [-2\pi,-\pi) \cup (\pi,2\pi] \\ 0, & \text{otherwise.} \end{cases}$$

The localized properties of the Shannon wavelet are apparent in the plot of ψ (Figure 1). The mother wavelet gives birth to a family of dilated and translated wavelets:

$$\left\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)\right\}_{j,k\in\mathbf{Z}}$$

where $\mathbf{Z} = \{0, \pm 1, \pm 2, \ldots\}$. From the time and frequency properties of ψ, it follows that $\psi_{j,k}$ will be well localized in time around $2^j k$ and effectively supported in the frequency domain on the set $[-2^{1-j}\pi,-2^{-j}\pi) \cup (2^{-j}\pi,2^{1-j}\pi]$.

The wavelet family $\{\psi_{j,k}\}$ constitute an orthonormal basis of $\mathscr{L}^2$ in which $\int \psi_{j,k}\psi_{j',k'} = \delta_j(j')\delta_k(k')$, where δ is the unit mass function (Mallat, 1989). As an orthonormal basis, any function $x \in \mathscr{L}^2$ can be represented as a superposition of the wavelet-basis functions:

$$x(t) = \sum_{j\in\mathbf{Z}}\sum_{k\in\mathbf{Z}} W_{j,k}^{(x)}\psi_{j,k}(t),$$

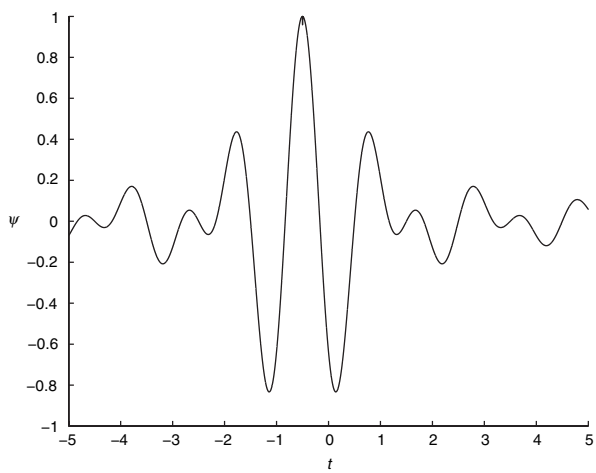

FIGURE 1. Shannon wavelet.

where:

$$W_{j,k}^{(x)} = \int x(t)\psi_{j,k}(t)\,\mathrm{d}t,$$

is the wavelet coefficient of $x(t)$ at scale j and translation k. The wavelet coefficients represent the projection of $x(t)$ onto the time-scale domain of the wavelet basis, $\{\psi_{j,k}\}$, and possess the frequency characteristics of $x(t)$ on the octave $[-2^{1-j}\pi,-2^{-j}\pi) \cup (2^{-j}\pi,2^{1-j}\pi]$ as found at the point in time $2^j k$. Hence, $W_{j,k}^{(x)}$ fully characterize $x(t)$ in the sense that the norm of $W_{j,k}^{(x)}$ equals the norm of $x(t)$, and if $W_{j,k}^{(x)}$ is topologically close to:

$$W_{j,k}^{(y)} = \int y(t)\psi_{j,k}(t)\mathrm{d}t, \quad \text{where } y(t) \in \mathcal{L}^2,$$

then $x(t)$ will be topologically close to $y(t)$.

We now project the LMSV model's linear offset representation of Equation (2) into the wavelet's time-scale space to produce the linear relationship:

$$W_{j,k}^{(y^*)} = W_{j,k}^{(\log \sigma^2)} + W_{j,k}^{(h)} + W_{j,k}^{(z)},$$

where $j = 1,\ldots,J = [\log_2 T]$, where $[\cdot]$ is the integer portion of the argument and $k = 1,\ldots,T/2^j$. The finite number of observations in y^* reduces the number of wavelet dilations to J and number of translations at each scale j to $T/2^j$. Note that, when the number of observations T is a power of 2 then the number of dilated and translated wavelet coefficients will be equal to $T - 1$. For convenience, we assume that T is always a power of 2, however, if T is not a power of 2 one simplly pads their data until it is by appending in reverse the observed data.

Without any loss of generality, we drop $W_{j,k}^{(\log \sigma^2)}$ from the wavelet representation of y^* and write $W_{j,k}^{(y^*)}$ as:

$$W_{j,k}^{(y^*)} = W_{j,k}^{(h)} + W_{j,k}^{(z)}. \tag{4}$$

This is possible because as a constant $\log \sigma^2$ wavelet coefficients, $W_{j,k}^{(\log \sigma^2)}$, will equal zero for every j and k. It follows that:

$$\log \sigma^2 \neq \sum_j \sum_k W_{j,k}^{(\log \sigma^2)} \psi_{j,k}(t), \quad \text{when } W_{j,k}^{(\log \sigma^2)} = 0.$$

However, by definition, an orthonormal basis must be able to represent the constant function $\log \sigma^2$. Wavelets accomplish this with a scaling function $\phi_J(t)$, where $\phi_J \in \mathcal{L}^2$ and satisfies the condition $\int \phi_J(t)\,\mathrm{d}t = 1$; i.e. ϕ_J, like ψ, has good local spatial and frequency properties except that it is a weighted moving average filter the effective frequency support of which is $[-\pi/2^J,\pi/2^J]$.[7] The scaling coefficient, $V_J^{(x)} = \int x(t)\phi_J(t)\,\mathrm{d}t$, is thus a weighted average of $x(t)$. It follows that:

$$\log \sigma^2 = V_J^{(\log \sigma^2)} \phi_J(t) + \sum_{j=1}^{J} \sum_{k=1}^{T/2^j} W_{j,k}^{(\log \sigma^2)} \psi_{j,k}(t)$$
$$= V_J^{(\log \sigma^2)} \phi_J(t) + 0.$$

Since the wavelet coefficients $\{W_{j,k}^{(y^*)}\}$ number only $T-1$, it is the addition of y^* scaling coefficient, $V_J^{(y^*)}$, that brings about the wavelet representation:

$$y_t^* = V_J^{(y^*)} \phi_J(t) + \sum_{j=1}^{J} \sum_{k=1}^{T/2^j} W_{j,k}^{(y^*)} \psi_{j,k}(t).$$

Jensen (2000) and McCoy and Walden (1996) both prove that the wavelet basis is a 'near' diagonalizing operator, where the wavelet coefficients of a fractionally integrated process are:

$$W^{(h)} \sim \mathcal{N}(\mathbf{0}, \Sigma_{W^{(h)}}), \tag{5}$$

where:

$$W^{(h)} = (W_{1,1}^{(h)}, W_{1,2}^{(h)}, \ldots, W_{1,T/2}^{(h)}, W_{2,1}^{(h)}, \ldots, W_{J-1,1}^{(h)}, W_{J-1,2}^{(h)}, W_{J,1}^{(h)})',$$

$$\mathbf{\Sigma}_{W^{(h)}} \approx \operatorname{diag}(\sigma_{1,1}, \sigma_{1,2}, \ldots, \sigma_{1,T/2}, \sigma_{2,1}, \ldots, \sigma_{J-1,1}, \sigma_{J-1,2}, \sigma_{J,1}),$$

and:

$$\sigma_{j,k} = \sigma_d^2 2^{2dj}, \quad j = 1, \ldots, J, \quad \text{and} \quad k = 1, \ldots, T/2^j,$$

with:

$$\sigma_d^2 = 2^{1-2d} \pi^{-2d} \sigma_\eta \frac{(1 - 2^{2d-1})}{(1 - 2d)}.$$

The frequency domain properties of ψ provide an intuitive understanding of this near independence result. By definition, the variance of a filtered process equals the energy (as measured by the area under the power spectrum) of the original process over the support of the filter. Hence, the variance of $W_{j,k}^{(h)}$ at a given scale j is proportional to the integral of $S(\omega)$ over the set $[-2^{1-j}\pi, -2^{-j}\pi) \cup (2^{-j}, 2^{1-j}\pi]$. More succinctly, since $S(\omega)$ is an even function one can integrate $S(\omega)$ over $(2^{-j}, 2^{1-j}\pi]$, multiply it by 2^j to normalize for the support length of $\hat{\psi}_{j,k}$ and obtain the wavelet variance, $\operatorname{Var} W_{j,k}^{(h)} = \sigma_d^2 2^{2dj}$.

The error in approximating $\Sigma_{W^{(h)}}$ with only the diagonal variance elements is a decreasing function of the wavelet's number of vanishing moments, M, where $M \geq 1$. Jensen (2000; Thm 2) shows that the correlation between $W_{j,k}^{(h)}$ and $W_{j',k'}^{(h)}$ decays over scale and translation at the exponential rate $\mathcal{O}(|2^{-j}k - 2^{-j'}k'|^{2(d-M)-1})$. This decorrelation is important for two reasons: first, although the elements of $W^{(h)}$ are not perfectly independent they are substantially less correlated than the original long-memory time series, h, and in light of the rate of decay in their

correlation independence is a reasonable assumption; second, the off-diagonal elements of $\Sigma_{W^{(h)}}$ are not known analytically, whereas $\text{var}(W_{j,k}^{(h)})$ are known in analytical form.[8]

By projecting y^* into the time-scale space of the wavelet basis, our objective of drawings from $\pi(\theta,h,s|y^*)$ has now changed to designing an MCMC simulator that draws from the LMSV model's wavelet domain posterior distribution, $\pi(\theta,W^{(h)},s|W^{(y^*)})$, where $\theta = (\sigma_d^2, d)$, and $W^{(y^*)}, W^{(h)}$ and s are all vectors of equal length having the same ordered indices as illustrated by $s = (s_{1,1},s_{1,2},\ldots,s_{1,T/2}, s_{2,1},\ldots,s_{J-1,1},s_{J,1})'$.

The near-independent, multivariate, Gaussian distribution of $W^{(h)}$ affords us an easy and efficient method of sampling the latent volatilities. With our wavelet representation of the LMSV we now sample $W^{(h)}|W^{(y^*)},s,\theta$ from:

$$\pi(W_{j,k}^{(h)}|W^{(y^*)},s,\theta) \propto \pi(W_{j,k}^{(h)}|\theta) f(W^{(y^*)}|W_{j,k}^{(h)},s,\theta),$$

where $f(W^{(y^*)}|W_{j,k}^{(h)},s,\theta)$ is the likelihood function and $\pi(W_{j,k}^{(h)}|\theta)$ is defined by Equation (5). We will detail this conditional distribution and its first and second moments in Section 3.4, but for now it should be clear how the near-independent nature of $W_{j,k}^{(h)}$ allows us to sidestep the difficult task of sampling the highly correlated h_t.[9]

3.2. *Mixture parameter selection*

As in the time-domain case, $\theta|W^{(y^*)},W^{(h)}$ and $W^{(h)}|W^{(y^*)},\theta\ W^{(z)}$ are intractable since both $W_{j,k}^{(y^*)} = W_{j,k}^{(h)} + W_{j,k}^{(z)}$ and $W_{j,k}^{(z)}$ are not normally distributed. Furthermore, unlike in the time domain where z_t distribution is known, the density $f(W_{j,k}^{(z)})$ is unknown, as are $W_{j,k}^{(z)}$ moments. Hence, we are unable to use Kim *et al.*'s (1998) method of setting the mixture parameters to match the first four moments of the mixture distribution with those of $f(W_{j,k}^{(z)})$. Instead, we adopt the semiparametric Bayesian approach of Tiwari *et al.* (1988), Chib and Hamilton (2002), and Hirano (2002), where the unknown distribution of the innovations are modelled non-parametrically by a Dirichlet process. In essence, we relax the normality assumption of ξ_t found in Equation (1) and model ξ_t as an innovation the distribution of which is unknown and its form unrestricted.[10]

Ferguson (1983) shows that any probability density function (pdf) can be arbitrarily approximated to a desired $\mathscr{L}^1$ level of accuracy by a countable mixture of independent normals. In the context of the unknown $f(W_{j,k}^{(z)})$ this implies:

$$f(W_{j,k}^{(z)}) = \sum_{i=1}^{\infty} \pi_i f_{\mathcal{N}}(W_{j,k}^{(z)}|\mu_i, v_i^2), \tag{6}$$

where $\sum_i \pi_i = 1$, $\pi_i \geq 0$, $f_{\mathcal{N}}(\cdot|\mu_i, v_i^2)$ is a normal pdf with mean μ_i and variance v_i^2, and $\mu_i \in \mathbb{R}$, and $v_i^2 > 0$, for $i \in \mathbf{Z}^+$. This representation of $f(W_{j,k}^{(z)})$ is equivalent to $W_{j,k}^{(z)}$ being conditionally independent and distributed:

$$W_{j,k}^{(z)}|\mu_{j,k}, v_{j,k}^2 \sim \mathcal{N}(\mu_{j,k}, v_{j,k}^2).$$

Denote the mixture distribution's unknown mean and variance parameters by the vector $\vartheta_{j,k} = (\mu_{j,k}, v_{j,k}^2)'$. To produce a Bayesian estimator of d we estimate $f(W_{j,k}^{(z)})$ by augmenting its unknowns with the other unknowns, $\theta, W^{(h)}$; i.e. we sample from $\theta, W^{(h)}, \vartheta | W^{(y^*)}$ where:

$$\vartheta = (\vartheta_{1,1}', \vartheta_{1,2}', \ldots, \vartheta_{1,T/2}', \vartheta_{2,1}', \ldots, \vartheta_{J-1,2}', \vartheta_{J,1}')',$$

has the same dual indexing as the vector s.

We assume *a priori* that the $\vartheta_{j,k}$ are independently distributed according to the unknown distribution function P; i.e.

$$f(W_{j,k}^{(z)}) = \int f_{\mathcal{N}}(W_{j,k}^{(z)}|\mu, v^2)\, \mathrm{d}P(\mu, v^2),$$

where P is the probability measure on the space, $\mathbb{R} \times \mathbb{R}^+$, with weight π_i on the points $\{\mu_i, v_i^2\}_{i \in \mathbf{Z}^+}$. In addition, we assume P is sampled from a dense Dirichlet process with mean distribution $P_0 = E[P]$ and precision a_0, where $a_0 > 0$. For now assume that a_0 is constant. Later we will relax this assumption and sample a_0 along with the other mixture unknowns.

From the definition of a Dirichlet process it follows that:

$$P \sim \mathcal{D}(a_0 P_0) \equiv \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \pi_{j,k}\, \delta_{\vartheta_{j,k}},$$

where $\{\pi_{j,k}\}$ and $\{\vartheta_{j,k}\}$ are independent, $\vartheta_{j,k} \overset{\text{i.i.d.}}{\sim} P_0$, and $\delta_{\vartheta_{j,k}}$ denotes the unit point mass at the argument $\vartheta_{j,k}$ (see Sethuraman and Tiwari, 1982; Ferguson 1983). To simplify the derivation of $\vartheta_{j,k}$ conditional distribution, we assume P_0 to be of the conjugate form:

$$v_{j,k}^2 \sim IG(\alpha, \beta) \quad \mu_{j,k}|v_{j,k}^2 \sim \mathcal{N}(\xi, \kappa v_{j,k}^2). \tag{7}$$

For a given realization of $W^{(z)}$, one can envision the mixture parameters, $\vartheta_{j,k}$, being independently sampled from P to produce the realization ϑ. Given the sampled ϑ, the observed $W_{j,k}^{(z)}$ are realizations from $\mathcal{N}(\mu_{j,k}, v_{j,k}^2)$. Since $P \sim \mathcal{D}(a_0 P_0)$ it follows that:

$$\vartheta_{j,k}|\vartheta^{(jk)}, W^{(z)} \sim \pi_{jk,0} P_{jk,0} + \sum_{(j,k) \neq (j',k')} \pi_{jk,j'k'}\, \delta_{\vartheta_{j,k}}(\vartheta_{j',k'}), \tag{8}$$

where $\vartheta^{(jk)}$ is the vector consisting of all the elements of ϑ other than $\vartheta_{j,k}$, and the probabilities of sampling from a particular component are:

$$\pi_{jk,0} \propto a_0 f_t(W_{j,k}^{(z)}|\xi, (1+\kappa)\beta/\alpha, \alpha), \tag{9}$$

$$\pi_{jk,j'k'} \propto f_{\mathcal{N}}(W_{j,k}^{(z)}|\mu_{j',k'}, v_{j',k'}^2), \quad j' \neq j, \quad k' \neq k, \tag{10}$$

where:

$$\pi_{jk,0} > 0, \quad \pi_{jk,j'k'} > 0, \quad \pi_{jk,0} + \sum_{j' \neq j} \sum_{k' \neq k} \pi_{jk,j'k'} = 1,$$

and f_t is the Student's t pdf:

$$f_t(W_{j,k}^{(z)} | \xi, (1+\kappa)\beta/\alpha, \alpha) = \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2)\sqrt{\alpha\pi(1+\kappa)\beta/\alpha}} \left(1 + \frac{1}{\alpha}\left(\frac{W_{j,k}^{(z)} - \xi}{\sqrt{(1+\kappa)\beta/\alpha}}\right)^2\right)^{-\frac{\alpha+1}{2}}$$

the mean of which is ξ, with scale factor $((1+\kappa)\beta/\alpha)^{1/2}$ and has α degrees of freedom (see Escobar 1994, Thm 1). From the priors in Equation (7), the posterior distribution $P_{jk,0} \equiv P_0 | W_{j,k}^{(z)}$ has the components:

$$v_{j,k}^2 | W_{j,k}^{(z)} \sim IG((1+\alpha), \beta_{j,k}), \quad \mu_{j,k} | W_{j,k}^{(z)}, v_{j,k}^2 \sim \mathcal{N}(\xi_{j,k}, \kappa v_{j,k}^2/(1+\kappa)), \tag{11}$$

with:

$$\beta_{j,k} = \beta + (W_{j,k}^{(z)} - \xi)^2/(1+\kappa) \quad \text{and} \quad \xi_{j,k} = (\xi + \kappa W_{j,k}^{(z)})/(1+\kappa).$$

Escobar (1994) and Escobar and West (1995) prove that draws from the posterior distribution $\vartheta | W^{(z)}$ can be made with a Gibbs sampler over the conditional distributions $\vartheta_{jk} | \vartheta^{(jk)}, W^{(z)}$ defined in Equation (8). A Gibbs sampler over Equation (8), however, can be inefficient since the finite number of different elements in ϑ can cause the sampler to become stuck at particular value of ϑ; i.e. too few draws being made from $P_{jk,0}$. We choose instead to use the more efficient algorithm of MacEachern (1992) and West *et al.* (1994), which produces draws that are equivalent to sampling from $\vartheta | W^{(z)}$ but by sampling from $\upsilon, s, K | W^{(z)}$, where $\upsilon = (\upsilon_1, \ldots, \upsilon_K)$ is a $2 \times K$ matrix consisting of the distinct elements of ϑ and $K \leq T - 1$.

Define the elements of the state vector s to be $s_{j,k} = i$, when $\vartheta_{j,k} = \upsilon_i$. In addition, let the matrix $\upsilon^{(jk)} = (\upsilon_1^{(jk)}, \ldots, \upsilon_{K^{(jk)}}^{(jk)})$, where $K^{(jk)} \leq T - 2$, be the distinct elements of $\vartheta^{(jk)}$ and let $n_i^{(jk)} = \#\{\vartheta_{j',k'} : \vartheta_{j',k'} = \upsilon_i^{(jk)}\}$, $i = 1, \ldots, K^{(jk)}$, represent the number of $\upsilon_i^{(jk)}$ elements in $\vartheta^{(jk)}$. Knowledge of υ, s and K is then equivalent to knowing ϑ, which allows Equation (8) to be rewritten as:

$$\vartheta_{j,k} | \vartheta^{(jk)}, W^{(z)} \sim \pi_{jk,0} P_{jk,0} + \sum_{i=1}^{K^{(jk)}} n_i^{(jk)} \pi_{jk,i} \, \delta_{\vartheta_{jk}}(\upsilon_i^{(jk)}), \tag{12}$$

where all the variables are defined in the same manner as in Equations (8)–(11) except with (10) becoming:

$$\pi_{jk,i} \propto f_{\mathcal{N}}(W_{j,k}^{(z)} | \mu_i, v_i^2). \tag{13}$$

Just as a given realization of $W^{(z)}$ can be viewed as the outcome of draws from $\vartheta_{j,k} \sim P_0$ and $W^{(z)}_{j,k}|\vartheta_{j,k} \sim \mathcal{N}(\mu_{j,k}, v^2_{j,k})$, ϑ can be thought of as resulting from a sequential draw of the elements of s from:

$$s_{j,k}|s_{1,1}, s_{1,2}, \ldots, s_{j,k-1} \sim \frac{a_0}{a_0 + n_{j,k} - 1}\,\delta_{s_{j,k}}(0) + \sum_{i=1}^{K_{j,k}} \frac{n_i}{a_0 + n_{j,k} - 1}\,\delta_{s_{j,k}}(i),$$

where $n_{j,k} = \#\{s_{j',k'} : j' \le j,\ k' < k\}$, $n_i = \#\{s_{j',k'} = i : j' \le j,\ k' < k\}$, and $K_{j,k}$ is the number of distinct elements in $(s_{1,1},\ldots,s_{j,k-1})$. If $s_{j,k} = 0$, then $\vartheta_{j,k}$ is sampled from P_0 and the matrix υ is increased by one column with a new mean and variance; i.e. $\upsilon_{K_{j,k+1}}|s_{j,k} = 0 \sim P_0$ and $K_{j,k+1} = K_{j,k} + 1$.

By Bayes theorem the posterior $\upsilon, s, K|W^{(z)}$ can be simulated by:

Step 1. Drawing $s_{j,k}$, $j = 1,\ldots,J$, $k = 1,\ldots,T/2^j$, from:

$$s_{j,k}|W^{(z)}, \vartheta^{(jk)}, K^{(jk)} \sim \pi_{jk,0}\,\delta_{s_{j,k}}(0) + \sum_{i=1}^{K^{(jk)}} n_i^{(jk)} \pi_{jk,i}\,\delta_{s_{j,k}}(i), \tag{14}$$

where if $s_{j,k} = 0$, υ_{K+1} is sampled from the posterior distribution, $P_{jk,0}$, defined in Equation (11) and K increases by 1.

Step 2. Given s and K, sample υ_i, $i = 1,\ldots,K$, from $\upsilon_i|W^{(z)}, s, K$ by:

(a) Drawing v_i^2 from $IG(\alpha + \#I_i, \beta_i)$, where $\#I_i$ denotes the cardinality of the set $I_i = \{j,k : s_{j,k} = i\}$,

$$\beta_i = \beta + \sum_{j,k \in I_i} \left(W_{j,k} - \bar{W}_i^{(z)}\right)^2 + \frac{\#I_i}{1 + \kappa \#I_i}\left(\xi - \bar{W}_i^{(z)}\right)^2,$$

and $\bar{W}_i^{(z)} = \sum_{j,k \in I_i} W_{j,k}^{(z)} / \#I_i$.

(b) Drawing μ_i from $\mathcal{N}(\xi_i, \kappa v_i^2/(1 + \kappa \#I_i))$, where

$$\xi_i = \left(\kappa \sum_{j,k \in I_i} W_{j,k}^{(z)} + \xi\right)/(1 + \#I_i \kappa).$$

In contrast to the inefficient Gibbs sampler over the conditional distribution of Equation (8), the $\upsilon, s, K|W^{(z)}$ sampler's superior mixing behaviour is apparent in Step 2 where each sweep is likely to produce new values of υ.

Instead of holding the Dirichlet process' precision parameter a_0 constant, Escobar and West (1995) suggest letting the data determine their value by sampling a_0 from $a_0|K$. Assuming the prior $\pi(a_0) = \Gamma(a_1, a_2)$, $a_1 > 0$, values of a_0 can be sampled from $a_0|K$ with the following Gibbs sampler:

Step 1. Draw $u \sim \text{Beta}(a_0 + 1, T - 1)$.

Step 2. Given u, sample a_0 from:

$$\pi_u \Gamma(a_1 + T - 1, a_2 - \ln u) + (1 - \pi_u)\Gamma(a_1 + K, a_2 + \ln u),$$

where π_u is defined by $\pi_u/(1 - \pi_u) = (a_1 + T - 2)/[(T - 1)(a_2 - \ln u)]$.

These two steps will be appended to the sampler of $\upsilon,s,K|W^{(z)}$ with the hyperparameters set equal to $a_1 = 2$, $a_2 = 0.5$, $\xi = 0$, $\kappa = 4/0.01$, $\alpha = 4$, and $\beta = 2$.[11]

3.3. *Priors*

To complete our Bayesian sampler we must specify the form of the prior $\pi(\theta)$. We first assume that all the unknown parameters are mutually independent. For d, the prior is set equal to $d = d^*/2$, where d^* is a random variable from a Beta distribution with the hyperparameters $d^{(1)}$ and $d^{(2)}$. It follows that the density is:

$$\pi(d) \propto (2d)^{d^{(1)}-1}(1-2d)^{d^{(2)}-1} \quad \text{where } d^{(1)}, d^{(2)} > 0, \tag{15}$$

which has an expected value of $0.5\ d^{(1)}/(d^{(1)} + d^{(2)})$.

The advantage of $\pi(d)$ is its ability to impose stationarity and long memory on h through its support, (0,1/2). This prior is also flexible enough to be an ignorant, uniform distribution on (0,1/2) when $d^{(1)} = d^{(2)} = 1$, or subjectively informative if we were to use historical data or institutional knowledge in setting the values of $d^{(1)}$ and $d^{(2)}$. In the simulations of Section 4 we use a non-informative prior by setting $d^{(1)} = d^{(2)} = 1$.

Because the value of σ_d^2 is not important in describing the long-memory characteristics of h, we assume our initial knowledge concerning it is accurately represented by the diffuse, inverse gamma distribution:

$$\pi(\sigma_d^2|v_0, \delta_0) \propto (\sigma_d^2)^{-(v_0+2)/2} \exp\left\{-\frac{\delta_0}{2\sigma_d^2}\right\}, \tag{16}$$

with hyperparameters equal to $\delta_0 = 2$ and $v_0 = 0.02$.

3.4. *Details of the MCMC algorithm*

To sample from the desired joint posterior distribution, $\theta,W^{(h)},s,\upsilon,K,a_0|W^{(y^*)}$, we design a hybrid MCMC sampler that combines the sampler of the mixture distribution parameters found in Section 2 with the Metropolis-Hastings (MH) algorithm (Metropolis *et al.*, 1953; Hastings, 1970) developed in this section. Our hybrid MCMC algorithm can be succinctly stated as follows:

Step 1. Initialize s,υ,K,a_0.
Step 2. Jointly sample $[\theta,W^{(h)}]$ from $\theta,W^{(h)}|W^{(y^*)},s,\upsilon$ by drawing,
  (a) θ from $W^{(h)}|W^{(y^*)},\theta,s,\upsilon$.
  (b) $W^{(h)}$ from $W^{(h)}|W^{(y^*)},\theta,\upsilon$
Step 3. Sample s,υ,K,a_0 from $s,\upsilon,K,a_0|W^{(y^*)},W^{(h)}$.

The MCMC algorithm begins in Step 1 with the initialization of the mixture distribution's indicator vector, s, parameter vector, υ, number of components, K, and the Dirichlet measure of precision, a_0. Using the method of composition, θ and $W^{(h)}$ are jointly sampled in Step 2 by first sampling θ in Step 2(a) from $\theta|W^{(y^*)},s,\upsilon$, where $W^{(h)}$ has been marginalized out and then drawing $W^{(h)}$ in Step 2(b) from $W^{(h)}|W^{(y^*)},s,\upsilon,\theta$. Chib and Carlin (1999) and Chib *et al.* (2002a, 2002b) find that a reduced blocking scheme where θ and $W^{(h)}$ are sampled as a single block reduces the serial correlation in the MCMC chain.[12] Intuitively, a reduced blocking scheme where draws are made from $\theta,W^{(h)}|W^{(y^*)},s,\upsilon$ eliminates the functional dependency that exists between $W^{(h)}$ and its parameters, θ, whereas, sampling separately from $\theta|W^{(y^*)},W^{(h)},s,\upsilon$, and $W^{(h)}|W^{(y^*)},\theta,s,\upsilon$, results in strongly correlated draws. Step 3, involves sampling s,υ,K and a_0 in the manner spelled out in Section 3.2.

The conditional distribution, $\theta|W^{(y^*)},s,\upsilon$, in Step 2(a) is proportional to:

$$\pi(\theta|W^{(y^*)},s,\upsilon) \propto \pi(\theta)f(W^{(y^*)}|s,\upsilon,\theta), \tag{17}$$

where from Equations (5) and (6) the density $f(W^{(y^*)}|s,\upsilon,\theta)$ equals:

$$f(W^{(y^*)}|s,\upsilon,\theta) = \prod_{j=1}^{J}\prod_{k=1}^{T/2^j}\frac{1}{\sqrt{2\pi(\sigma^2_{s_{j,k}}+\sigma^2_d 2^{2dj})}}\exp\left\{-\frac{(W^{(y^*)}_{j,k}-\mu_{s_{j,k}})^2}{2(\sigma^2_{s_{j,k}}+\sigma^2_d 2^{2dj})}\right\}. \tag{18}$$

Equation (17) is a non-standard distribution and hence, draws from $\theta|W^{(y^*)},s,\upsilon$ cannot be made with existing random number generators. We choose to employ the tailored Metropolis–Hasting algorithm to sample from this intractable density (see Chib and Greenberg 1995, 1998 for a detailed account of this sampler). With the tailored MH sampler we make a candidate draw of $\theta|W^{(y^*)},s,\upsilon$ by sampling from the multivariate Student's t density $f_t(\cdot|m,(\eta-2)V,\eta)$, whose mean vector is m, with scale matrix, $(\eta-2)V$, and η degrees of freedom. In practice we set the Student's t mean and covariance equal to:

$$m = \arg\max_{\theta} \log\, \pi(\theta)f(W^{(y^*)}|s,\upsilon,\theta),$$
$$V = [-\partial^2 \log\, \pi(\theta)f(W^{(y^*)}|s,\upsilon,\theta)/(\partial\theta\,\partial\theta')\big|_{\theta=m}]^{-1}$$

the values of which are obtained with the quasi-Newton maximization algorithm developed by Broyden, Fletcher, Goldfarb, Shanno (see Fletcher, 1987) initialized at the value of θ from the MCMC sampler's previous sweep. The degrees of freedom is set to $\eta = 10$.

Denote the candidate draw from the Student's t distribution as θ'. The draw, θ', is accepted as a realization from $\theta|W^{(y^*)},s,\upsilon$ with MH probability:

$$\alpha(\theta,\theta'|W^{(y^*)},s,\upsilon) = \min\left\{\frac{\pi(\theta')f(W^{(y^*)}|s,\upsilon,\theta')}{\pi(\theta)f(W^{(y^*)}|s,\upsilon,\theta)}\frac{f_t(\theta|m,(\eta-2)V,\eta)}{f_t(\theta'|m,(\eta-2)V,\eta)},1\right\},$$

where θ is the value of the parameter from the previous sweep. In other words, θ' will be accepted with probability $\alpha(\theta,\theta'|W^{(y^*)},s,\upsilon)$ as a realization from $\theta|W^{(y^*)},s,\upsilon$, or conversely, θ will be kept with probability $1-\alpha(\theta,\theta'|W^{(y^*)},s,\upsilon)$ as the sampler's current realization.

Because of the wavelets approximate independent nature, draws in Step 2(b) from the multivariate distribution $W^{(h)}|W^{(y^*)},\theta,s,\upsilon$ can be sampled individually from the tractable univariate distribution:

$$\pi(W_{j,k}^{(h)}|W^{(y^*)},\theta,s,\upsilon) \propto \pi(W_{j,k}^{(h)}|\theta) f(W^{(y^*)}|W_{j,k}^{(h)},\theta,s,\upsilon),$$

where $\pi(W_{j,k}^{(h)}|\theta) \sim \mathcal{N}(0,\sigma_d^2 2^{2dj})$ and the likelihood function equals:

$$f(W^{(y^*)}|W_{j,k}^{(h)},\theta,s,\upsilon) = \prod_{j'=1}^{J}\prod_{k'=1}^{T/2^{j'}} \frac{1}{\sqrt{2\pi(\sigma_d^{2dj'}+\sigma_{s_{j'k'}}^2)}} \times \exp\left\{-\frac{\left(W_{j',k'}^{(y^*)}-W_{j,k}^{(h)}-\mu_{s_{j'k'}}\right)^2}{2(\sigma_d^2 2^{2dj'}+\sigma_{s_{j'k'}}^2)}\right\}.$$

By completing the square on the kernel of $f(W^{(y^*)}|W_{j,k}^{(h)},\theta,s,\upsilon)$ and multiplying it by $f_{\mathcal{N}}(W_{j,k}^{(h)}|0,\sigma_d^2 2^{2dj})$, we find the conditional latent wavelet coefficients are approximately independent and distributed:

$$W_{j,k}^{(h)}|W^{(y^*)},\theta,s,\upsilon \sim \mathcal{N}(\bar{W}_{j,k},v_{j,k}^2),$$

where:[13]

$$\bar{W}_{j,k} = \left(\frac{(\bar{\sigma}_{j,k}^2)^{-1}}{(\bar{\sigma}_{j,k}^2)^{-1}+(\sigma_d^2 2^{2dj})^{-1}}\right)\bar{W}_{j,k}^{(y^*)}, \quad v_{jk}^2 = \frac{1}{(\bar{\sigma}_{j,k}^2)^{-1}+(\sigma_d^2 2^{2dj})^{-1}},$$

with:

$$\bar{W}_{j,k}^{(y^*)} = W_{j,k}^{(y^*)} - \mu_{s_{j,k}}, \quad \text{and} \quad \bar{\sigma}_{j,k}^2 = \sigma_{s_{j,k}}^2 + \sigma_d^2 2^{2dj}.$$

4. EXAMPLES

In this section, we report the results from our MCMC algorithm when it is applied to simulated and daily stock return data. For both the artificially generated and real stock return data we use the Daubechies (1992), (p. 198) least asymmetric wavelet that has $M = 4$ vanishing moments; i.e., $\int t^n \psi(t)\, dt = 0$, for $n = 0,1,2,3$, to compute the log-squared returns wavelet coefficients. A number of issues were taken into consideration in choosing this wavelet; i.e. issues like boundary effects

and possible statistical artifacts caused by a wavelet with a large number of vanishing moments. We choose the Daubechies least asymmetric wavelet with four vanishing moments because it favourably addresses each of the above points while possessing the extraordinary theoretical property of diagonalizing a long-memory process' covariance matrix (Whitcher, 1998; Percival and Walden, 2000, pp. 346–349).

4.1. *Simulated LMSV*

Using the programming language OX (Doornik, 2001), we generate and estimate LMSV models comprised of $T = 4096$ observations for the array of long-memory parameter values, $d = 0.1,0.25,0.4$.[14] Each simulated model's innovation term, η_t, is Gaussian white noise with variance $\sigma_\eta^2 = 0.1$. We also let the instantaneous volatility, $\sigma = 1$.

The MCMC algorithm initial starting values are $d_0 = \sigma_{d,0}^2 = 0.333$, $\mu_{j,k} = 0$ and $\sigma_{j,k}^2 = 2.0$, for $j = 1,\ldots,J$ and $k = 1,\ldots,T/2^j$, for the simulated LMSV data cases and the empirical data as well. We used other starting values and found no distinguishable difference in the MCMC results from those reported here. We burn-in the MCMC algorithm to allow it to converge to the target posterior distribution by discarding the first 1000 of 6000 Gibbs draws.

To determine if our sampler has converged to the target posterior density and is making uncorrelated draws from the posterior distribution, we measure the observed serial correlation in the draws of the unknowns with its inefficiency measure:

$$1 + \frac{2N}{N-1} \sum_{\tau=1}^{L} K\left(\frac{\tau}{L}\right) \rho(\tau),$$

where $K(\cdot)$ is Parzen's filter (see Percival and Walden, 1998, p. 265), $\rho(\cdot)$ is the sample autocorrelation function of the drawn parameter, N is the number of draws ($N = 5000$), and L is the largest lag at which the autocorrelation function is computed (In the simulations below we find that the MCMC draws correlogram decays so rapidly that $L = 100$ is adequate.).

The inefficiency measure quantifies the loss associated with using the correlated draws from the sampler, as opposed to hypothetically independent draws, to compute the posterior mean. We also use the inefficiency measure to compute the posterior mean's numerical standard error (NSE). A Bayesian parameter estimate NSE is simply the square root of the product between the inefficiency measure and the sample variance of the draws (Geweke, 1992).

Table I contains the results from our MCMC algorithm for the simulated LMSV models. In the first five columns of Table I the true value of d, and the posterior's sample mean, numerical standard error, inefficiency measure (Ineff.), and 5th and 95th percentiles (in parenthesis) for d are listed. The MH

TABLE I

THE SEMIPARAMETRIC BAYESIAN (GIBBS SAMPLE OF 6000 DRAWS WITH THE FIRST 1000 DRAWS DISCARDED), AND GPH (USING $[T^{1/2}]$ FREQUENCIES) ESTIMATOR AND THE PARAMETRIC QMLE ESTIMATOR OF THE LONG-MEMORY PARAMETER FOR SIMULATED LMSV DATA $[y_t = \exp\{h_t/2\}\xi_t, (1-L)^d h_t = \eta_t, t = 1, \ldots, 4096, \eta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,01)]$

	Bayesian							
d	Mean	NSE	Ineff.	90% Conf.	$\alpha(\theta, \theta')$	GPH	90% Conf.	QMLE
0.1	0.1145	0.1086	3.2201	0.0109, 0.2341	0.6878	0.0714	−0.0898, 0.2326	0.0323
0.25	0.2130	0.0854	2.7481	0.1037, 0.3063	0.8118	0.0826	−0.0786, 0.2438	0.0932
0.4	0.3092	0.0886	4.8751	0.2286, 0.3836	0.8410	0.1788	0.0176, 0.3400	0.2626

Notes: NSE is the numerical standard error, Ineff. is the inefficiency factor, 90% Conf. for the Bayesian estimates denotes the 5th and 95th percentiles of d posterior distribution, and the GPH 90% confidence interval is computed with the log-periodogram least-squares regressors standard errors, where the errors are assumed to be independent with mean zero and variance $\pi^2/6$. The asymptotic properties of the QMLE for the LMSV model are unknown hence its 90% confidence interval is not reported.

probability of acceptances, $\alpha(\theta,\theta')$, for the draw $\theta|W^{(y^*)},s,\upsilon$ are tabulated in the sixth column.

Table I also reports the the popular Geweke and Porter-Hudak (1983) estimate of d and its 90% confidence interval in the seventh and eighth columns using $[\sqrt{T}]$ frequencies. Like our Bayesian estimator, the GPH is a semiparametric estimator of d but is based on the log-periodogram regression. Because $\log y_t^2$ is not Gaussian, the asymptotic properties found by Robinson (1995) for the GPH estimator do not hold under the assumptions of the LMSV model. However, Deo and Hurvich (2001) find the GPH estimator of the LMSV model to be downwardly biased with an empirical standard deviation closer to the standard errors associated with regression theory than to its theoretical asymptotic standard errors. The values in the parenthesis found in Table I for the GPH estimates of d are its 90% confidence interval calculated with the regression's standard error.

The ninth column of Table I contains the estimates of d using Breidt *et al.*'s (1998) quasi-maximum likelihood estimator (QMLE). Unlike our Bayesian estimator and the GPH estimator, the QMLE is a fully parametric estimator of d that requires specification of the order of short-memory parameters. With the model fully specified, the QMLE is computed by maximizing the approximate frequency domain likelihood function under the assumption that y^* from Equation (2) is normally distributed. The QMLE of the LMSV has been shown to be strongly consistent (Breidt *et al.*, 1998), but its other asymptotic properties are not yet known. As a result we do not report a 90% confidence interval for the QMLE.

The inefficiency measures found in the fourth column of Table I indicate that our MCMC sampler is mixing well and producing near-independent draws of d from the posterior distribution; the inefficiency measures are all < 5. Thus, a draw of 5000 from our MCMC algorithm is sufficient for the Monte Carlo standard error of the posterior mean to be no more than one-tenth of 1%. These small

inefficiency measures are also evidence to the fact that integrating out the latent wavelet coefficients from the draw of θ, sampling θ with the tailored MH algorithm, drawing the latent volatilities from their wavelet domain, and using MacEachern (1992) and West *et al.* (1994) efficient sampler of the mixture parameters have led to a very efficient algorithm.

Further evidence of efficiency and the important role the tailored MH sampler plays in the MCMC algorithm is found in the values of the MH probability of acceptances, $\alpha(\theta,\theta')$. In Table I, all of the MH acceptance rates are $> 68\%$. Hence, the draws of θ are not getting stuck in a particular region of the conditional distribution, $\theta|W^{(v^*)},s,\upsilon$. Instead, the MH sampler is making uncorrelated draws from across the unknown distribution.

We find the Bayesian point estimate of the fractional order of integration in Table I when d is equal to 0.1 and 0.25 to be of a desirable nature. Each posterior distribution is concentrated around the true value of $d = 0.1,0.25$ as is evident in the posterior's mean being no more than 0.04 different from the true d and the NSE no larger than 0.11. The only noticeable difference is in the smaller MH acceptance rate when $d = 0.1$. This is understandable since in this case d is close to the zero lower bound of $\pi(d)$ and the likelihood of drawing a candidate value less than zero is greater when the LMSV models has a small-valued d.

When the true $d = 0.4$, the Bayesian estimate of the long-memory parameter is 0.3. By itself, this point estimate is not troubling, but combined with the estimate's small numerical standard error and its tight 90% Bayesian confidence interval, the true value of d fails to fall within the confidence interval (one should note, however, that the Bayesian estimator's 95th percentile of 0.3836 is close to capturing the true d). The two other Bayesian confidence intervals in Table I for $d = 0.1,0.25$ are also tight, but because their Bayesian point estimates do not underestimate the true d in these cases, their confidence intervals do capture the true fractional differencing parameter.

In comparison to the Bayesian estimator, both the GPH and QMLE estimates of d are substantially smaller than both the Bayesian estimates and the true long-memory parameter value. The GPH's standard error is also considerably larger than the Bayesian estimators. Finding that the GPH underestimates the true value of d is not surprising, given Deo and Hurvich's (2001) theoretical and empirical findings that the GPH estimator, on average, underestimates the degree of long-memory present in volatility. However, because of this underestimation, the GPH estimators 90% confidence interval fails to capture the true value of d in two of the three experiments ($d = 0.25,0.4$).

4.2. *Frequentist comparison*

An important area of research in Bayesian statistics is the construction of objective-ignorant priors where the chosen prior distribution is non-informative about the parameter's value (see Berger and Bernardo, 1992, and the references therein). One measure of a prior's level of objective ignorance is the frequentist

volume of its posterior's Bayesian confidence interval. Welch and Peers (1963) show that the frequentist volume of the Bayesian confidence interval will be close to the nominal level used to construct the confidence interval, if the prior is objectively ignorant. By conducting a Monte Carlo experiment and measuring the frequency at which the posterior distribution's 90% confidence interval captures the true value of d, we will be able to compare the objective ignorance of the prior.

In our Monte Carlo experiment we generate 100 stochastic volatility series where the latent volatility process h_t is one of three fractionally integrated models:

$$(1-L)^d h_t = \eta_t, \quad t = 1, \ldots, T,$$

with $d = 0.1,0.25,0.4$, $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.1)$, and $T = 1024$, 2048. For each LMSV model we estimate d for the 100 simulated series with the wavelet Bayesian estimator using the priors described in Section 3.3, the GPH estimator, and the QMLE. A 90% confidence interval, $(\hat{d}_{0.05}, \hat{d}_{0.95})$, is then constructed for the two semiparametric estimators.[15] The Bayesian confidence interval is found by selecting the simulated posterior's 5th and 95th quantiles of the drawn ds, whereas the GPH's confidence interval equal $\hat{d}_{\text{GPH}} \pm 1.64\sigma_{\text{GPH}}$, where σ_{GPH} is the standard deviation for the estimated slope in the GPH log-linear periodogram frequency regression under the assumption that the residuals are independent with mean zero and variance $\pi^2/6$.

Each of the three long-memory parameter estimator's empirical bias and MSE from the Monte Carlo experiments are reported in Table II, along with the empirical frequency at which the true d falls outside the Bayesian and GPH estimators 90% confidence intervals, i.e. $P(d < \hat{d}_{0.05})$ and $P(d > \hat{d}_{0.95})$. Both semiparametric estimators are downward biased, but when $d = 0.1,0.25$ the Bayesian estimator is significantly less biased than the GPH estimator. When $d = 0.4$, the bias of the two semiparametric estimators are nearly the same. We also find the GPH estimator's MSE to be significantly larger than the MSE of the Bayesian estimator for both small and large values of d. In slightly more than half of our experiments, the QMLE is also biased downward but of a smaller magnitude than the semiparametric Bayesian estimator in every case except for when $d = 0.1$. The empirical MSE of the QMLE, however, is larger than the Bayesian estimators in all six experiments.

There is a significant difference in the coverage of the two semiparametric estimators confidence intervals. The downward bias and the tightness of the posterior distribution causes the 90% Bayesian confidence interval to either over capture the true d (the case when $d = 0.1$), or the 90th percentile of the posterior distribution too often falls below the true d as is the case when $d = 0.25,0.4$. The GPH estimators' confidence intervals also suffer from a downward bias, but, because the GPH estimator has a larger standard error its confidence interval captures the true value of d more often than suggested by its asymptotic properties.

In the experiment where $d = 0.1$, the coverage of the Bayesian confidence intervals suggest that the uniform prior over the interval (0,1/2) for $\pi(d)$ is too

TABLE II

MONTE CARLO RESULTS FROM AN EXPERIMENT CONSISTING OF 100 LONG-MEMORY STOCHASTIC VOLATILITY MODELS WHERE THE LATENT VOLATILITY PROCESS IS FRACTIONAL INTEGRATED [$y_t = \exp\{h_t/2\}\xi_t,\ (1-L)^d h_t = \eta_t,\ t = 1,\ldots,T,\ \eta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,0.1)$].

		Bayesian				GPH				QMLE	
T	d	Bias	MSE	$P(d < \hat{d}_{0.05})$	$P(d > \hat{d}_{0.95})$	Bias	MSE	$P(d < \hat{d}_{0.05})$	$P(d > \hat{d}_{0.95})$	Bias	MSE
1024	0.1	0.0001	0.001	0	0	−0.098	0.16	0	0.01	−0.0463	0.0330
	0.25	−0.0980	0.012	0	0.28	−0.170	0.19	0	0.01	0.0063	0.1169
	0.4	−0.1800	0.034	0	0.85	−0.130	0.15	0	0.01	−0.0307	0.0559
2048	0.1	−0.0014	0.001	0	0	−0.067	0.12	0	0.01	−0.1233	1.0981
	0.25	−0.0860	0.010	0	0.32	−0.130	0.14	0	0.02	0.0132	0.1044
	0.4	−0.1100	0.014	0	0.76	−0.110	0.13	0	0.04	−0.0197	0.0175

Notes: The MCMC simulator for the wavelet, Bayesian estimator burns-in for 1000 draws and keeps the next 5000 draws. $P(d < \hat{d}_{0.05})$ and $P(d > \hat{d}_{0.95})$ are respectively the Monte Carlo probabilities that the true value of d was below or above the estimators 90% confidence interval. Empirical frequencies of d falling outside the QMLE's confidence interval are not possible since the QMLE asymptotic distribution is not known under the LMSV model.

informative. However, when this finding is contrasted with the number of times the Bayesian confidence intervals captures $d = 0.25,0.4$, the uniform prior appears to be too ignorant. If one knew before hand that d is large, better inference could be obtained by using this information in the prior of d. Such a prior would not only help the Bayesian confidence interval capture the true parameter value, but it would also help to reduce the downward bias in the Bayesian estimator.

4.3. *Stock returns*

Over the past few years, the presence of long memory in the volatility of equity returns and foreign exchange rates has come to dominate the literature on temporal dependencies in financial volatility. These findings of strong persistent volatilities have been discovered in both daily financial data measured over long time horizons, and high-frequency, tick-by-tick data recorded over shorter time spans. Up to this point, the empirical work on long-memory dependence in volatility has always involved the frequentist-based GPH or QMLE estimators.

For example, Bollerslev and Jubinski (1999) and Ray and Tsay (2000) have investigated the long-memory behaviour in the volatility of the AA daily stock returns. Bollerslev and Jubinski (1999) use Robinson's (1995) bivariate version of the GPH estimator to estimate the long-memory parameters of absolute returns and volume, whereas Ray and Tsay (2000) apply both the univariate GPH estimator and Breidt *et al.* (1998) QMLE to log-squared returns. Each find evidence of strong persistence in the volatility of the AA daily stock returns with a long-memory parameter estimate of approximately 0.35 and an asymptotic standard error close to 0.06.

In this subsection, we report the results of our MCMC algorithm as applied to the daily stock returns of AA. The return data we use is from the Center for Research in Security Prices database. The continuously compounded returns are corrected for the effects of stock splits and dividends and consist of 8938 daily observations from 3 July 1962, to 31 December 1997. Like the fast Fourier transform, the fast wavelet transform algorithm requires the sample size to be an integer power of 2 and treats the time series as if it were circular when calculating coefficients near the beginning and end of the data set. To apply the fast wavelet transform to the entire data set, we draw on the circular assumption of the transform and append in reverse the daily stock returns to the original data set until we obtain $2^{14} = 16{,}384$ observations.[16]

In Table III we report the same summary statistics for our MCMC simulator along with the GPH and QMLE estimators for the volatility of AA daily returns. We also plot in Figure 2 the 5000 posterior draws of d, their histogram, and the first 20 lags of its correlogram.[17] As we saw with the simulated data, the inefficiency factor in Table III suggests that our sampler is once again efficient.

TABLE III

BAYESIAN ESTIMATION OF A LONG-MEMORY STOCHASTIC VOLATILITY MODEL FOR ALUMINUM CORPORATION OF AMERICA DAILY COMPOUNDED RETURNS FROM 3 JULY 1962 TO 31 DECEMBER 1997 ($T = 8938$).

	Bayesian estimate					
Para.	Mean	NSE	Ineff.	$\alpha(\theta,\theta')$	GPH	QMLE
d	0.3481 (0.3043, 0.3903)	0.0952	13.1960	0.8918	0.3598 (0.2416, 0.4781)	0.4799

Note: The values in parenthesis are the 90% confidence intervals for the estimated d.

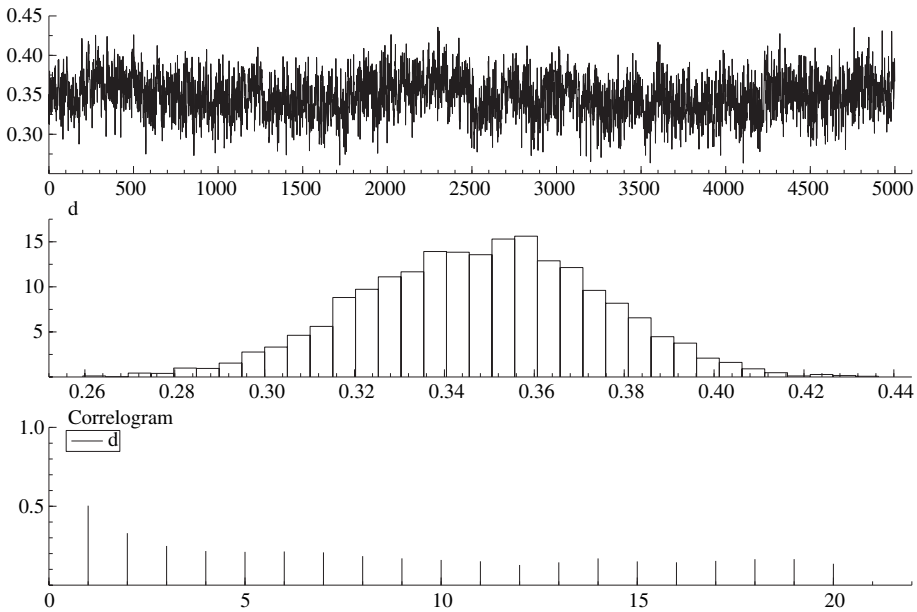

FIGURE 2. The 5000 MCMC draws of d, their histogram and correlogram for the Aluminum Corporation of America's daily compounded returns from 3 July 1962 to 31 December 1997 ($T = 8938$).

The sampler is mixing well as is evident in both d's inefficiency factor being <14 and in Figure 2 correlogram decaying rapidly to zero.

In terms of the posterior distribution of long-memory parameter, the estimated posterior density of d is concentrated around the posterior mean 0.3481 with a tight 90% Bayesian confidence interval of (0.3043,0.3903). These long-memory values all produce the strong persistent behaviour commonly found in the volatility of financial data. In comparison with the point estimates of d reported by Bollerslev and Jubinski (1999) and Ray and Tsay (2000), our estimate is similar to these earlier estimates of d. Ray and Tsay (2000) produce a slightly larger long-memory parameter estimate of 0.355, while Bollerslev and Jubinski (1999) report

a d of 0.379. Both of these estimates fall well within our posterior's 95% confidence interval.

5. CONCLUSION

In this paper, we have described how to quickly and efficiently sample from the posterior distribution of long-memory stochastic volatility model parameter with a MCMC. The key element to our MCMC sampler is the independent distribution of the latent variables afforded to us by the long-memory volatility's wavelet coefficients. This independence in the latent volatility's wavelet coefficients allows us to replace Kim *et al.* (1998) simulation smoother with a simple draw of the latent wavelet coefficients from a fully conditional, multivariate, independent normal distribution. The simulations conducted in this paper show that the proposed MCMC sampler is an appealing alternative to existing frequency based estimators of fractionally integrated models of long-memory stochastic volatility from which economists can garner confidence in using it to make inference about the degree of persistence inherent in volatility.

ACKNOWLEDGEMENTS

I would like to thank Sid Chib, Ed Greenberg, Keisuke Hirano, and Val Lambson; the participants at the Midwest Econometric Group Meetings held at the University of Chicago, the Symposium on Nonlinear Dynamics and Econometrics held at the Federal Reserve Bank of Atlanta, and the seminars at the University of California – Davis, the University of Technology – Sydney, Monash University, the University of Auckland, Brigham Young University, and Washington University for their constructive comments; and the Ira and Mary Lou Fulton Supercomputing Center at Brigham Young University for the use of its computing resources. I would also like to express my appreciation to the two anonymous referees for their helpful comments and suggestions. Part of this paper was completed while the author was a Visiting Fellow at the University of Technology in Sydney. Grant funding from the College of Family, Home, and Social Science at Brigham Young University enabled Scott Murdock to provide invaluable research assistance.

NOTES

1. So (2002) designs a more computationally taxing sampler than ours where artificial disturbances are used in a simulation smoother of the latent volatilities.

2. To enrich the short- and medium-run behavior of volatility we can add short-memory parameters to h_t and obtain Breidt *et al.* (1998) autoregressive, fractionally integrated, moving average, stochastic volatility model (ARFIMASV). We are currently involved in extending this paper's wavelet-based algorithm to ARFIMASV models and to the calculation of its likelihood function. Using Bayes factor, this will allow us to test between long- and short-memory models of volatility.
3. Meyer and Yu (2000) use the single-step sampler found in the BUGS software to sample latent volatility when leverage effects are present in a first-order autoregressive stochastic volatility model, whereas, Chan *et al.* (2003) design a multiple block sampler of a multivariate stochastic volatility model with leverage effects. It is unlikely that our efficient reduced-block sampler of long-memory volatility and its use of the linear offset method of Kim *et al.* (1998) and Chib *et al.* (2002a) can address leverage effects. Instead, a multiple block sampler seems to be required.
4. For a review of the work performed in economics with long-memory processes see Baillie (1996).
5. For an extensive introduction to wavelet analysis from the vantage point of time series analysis see Percival and Walden (2000), or see Hernández and Weiss (1996) for a theoretical viewpoint.
6. The admissibility condition is only a necessary condition for the existence of a wavelet. More stringent rates of decay can be required of $\hat{\psi}$ in theory and practice.
7. The scaling function corresponding to the Shannon wavelet is $\phi_J(t) = 2^J \sin(\pi 2^{-J}t)/(\pi 2^{-J}t)$, whose $\text{supp}(\hat{\phi}_J) = [-\pi/2^J, \pi/2^J]$.
8. One possible, but very taxing, way of computing the off-diagonal elements of $\Sigma_{W^{(h)}}$ is to use the analytical formula of a long-memory process autocovariance function to construct the autocovariance matrix. The entire wavelet coefficients covariance matrix is then equal to the tensor product of the wavelet transform applied to the long-memory process' autocovariance matrix; i.e. $\Sigma_{W^{(h)}}$ is equal to the projection of the autocovariance matrix into the wavelet domain by a two-dimensional wavelet transform.
9. An alternative idea for a sampler of $W^{(h)}$ is to use the approximate independent prior of $W^{(h)}$ as an auxiliary model to produce candidate draws. These draws would then be accepted or rejected based on a Metropolis-Hastings probability calculated with $W^{(h)}$ exact prior. This approach, however, is not feasible because $\Sigma_{W^{(h)}}$ does not have an analytical form and projecting the covariance matrix of h into the wavelet domain is to costly given the large number of sweeps the sampler requires.
10. Mahieu and Schotman (1998) allow the distribution of ξ_t in a first-order autoregressive stochastic volatility model to differ from normality. They model the distribution of z_t as a mixture of normals but choose to fix the number of components (to three) along with their probabilities. The means and variances of the mixture components are then estimated along with the model's other unknown parameters.

11. Hirano (1998) find that these values provide a prior of $f(W_{j,k}^{(z)})$ that contains a broad class of densities with weight placed on those that are relatively smooth.
12. Liu *et al.* (1994) were the first to show that grouping together highly dependent unknowns would increase the convergence rate of the sampler.
13. The mean and variance of $\pi(W_{j,k}^{(h)}|W^{(y^*)},\theta,s,\upsilon)$ have a nice intuitive appeal. $\bar{W}_{j,k}$ is a weighted average of $W_{j,k}^{(h)}$ prior's mean (which is zero) and the observed data reduced by the mixture's mean, $\bar{W}_{(j,k)}^{(y^*)}$, where the weight is proportional to the precision of the data, $(\bar{\sigma}_{j,k}^2)^{-1}/[(\bar{\sigma}_{j,k}^2)^{-1}+(\sigma_d 2^{2dj})^{-1}]$. The variance, v_{jk}^2, is the inverse of the linear combination of the prior's precision, $(\sigma_d 2^{2dj})^{-1}$, and the data's precision, $(\bar{\sigma}_{j,k}^2)^{-1}$.
14. We use the algorithm of Hosking (1984) to generate the latent fractionally integrated processes.
15. Inference cannot be made with the QMLE since its asymptotic properties are not known.
16. We also tried purging the first 746 observations of the original dataset to obtain $2^{13}=8192$ observations, and found our estimate of d to be nearly the same as those found using the method described in the body of the paper.
17. Although not reported we find the histogram and correlogram to be very similar for the simulated LMSV data when $d=0.4$ and $T=4096$.

REFERENCES

Adenstedt, R. K. (1974) On large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095–107.

Andersen, T. and Bollerslev, T. (1997) Heterogeneous information arrivals and return volatility dynamics: uncovering the long run in high frequency returns. *Journal of Finance* 52, 975–1005.

Baillie, R. T. (1996) Long-memory processes and fractional integration in econometrics. *Journal of Econometrics* 73, 5–59.

Beran, J. (1994) *Statistics for Long-memory Processes*. New York: Chapman & Hall.

Berger, J. O. and J. M. Bernardo (1992) On the development of reference priors. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, J. O. Dawd, A. F. M. Smith). New York, NY: Oxford University Press, pp. 35–60.

Bollerslev, T. and Jubinski, D. (1999) Equity trading volume and volatility: latent information arrivals and common long-run dependencies. *Journal of Business and Economic Statistics* 17, 9–21.

Breidt, F. J., Crato, N. and de Lima, P. (1998) On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics* 83, 325–48.

Chan, N. H. and Palma, W. (1998) State space modeling of long-memory processes. *Annals of Statistics* 26, 719–40.

Chan, D., Kohn, R. and Kirby, C. (2003) Multivariate stochastic volatility with leverage. Working Paper, Australian Graduate School of Management, University of New South Wales.

Chib, S. (2001) Markov chain Monte Carlo methods: computation and inference. In *Handbook of Econometrics*, Vol. 5, (eds J. J. Heckman and E. Leamer). Amsterdam: Elsevier Science, pp. 3569–649.

Chib, S. and Carlin, B. P. (1999) On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing* 9, 17–26.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings Algorithm. *American Statistician* 49, 327–35.

Chib, S. and Greenberg, E. (1996) Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409–31.

CHIB, S. and GREENBERG, E. (1998) Analysis of multivariate probit models. *Biometrika* 85, 347–61.
CHIB, S. and HAMILTON, B. (2002) Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* 110, 67–79.
CHIB, S., NARDARI, F. and SHEPHARD, N. (2002a) Markov chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* 108, 281–316.
CHIB, S., NARDARI, F. and SHEPHARD, N. (2002b) Analysis of high dimensional multivariate stochastic volatility models, unpublished manuscript, Olin School of Business, John M. Washington University.
DAUBECHIES, I. (1992) *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.
DEO, R. S. and HURVICH, C. M. (2001) On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686–710.
DIJKERMAN, R. W. and MAZUMDAR, R. R. (1994) On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Transaction of Information Theory* 40, 1609–12.
DOORNIK, D. F. (2001) *Ox: An Object-Oriented Matrix Language* (4th edn). London: Timberlake Consultants Press.
ESCOBAR, M. D. (1994) Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* 89, 268–77.
ESCOBAR, M. D. and WEST, M. (1995) Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90, 577–88.
FERGUSON, T. S. (1983) Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday* (eds H. RIZVI and J. RUSTAGI). New York: Academic Press, pp. 287–302.
FLETCHER, R. (1987) *Practical Methods of Optimization* (2nd edn). New York: Wiley.
FULLER, W. A. (1996) *Introduction to Statistical Time Series* (2nd edn). New York: Wiley.
GELFAND, A. E. and SMITH, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
GEWEKE, J. (1992) Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4* (eds J. M. BERNARDO, J. O. BERGER, J. O. DAWD, A. F. M. SMITH). New York: Oxford University Press, 169–93.
GEWEKE, J. and PORTER-HUDAK, S. (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221–38.
GHYSELS, E., HARVEY, A. C. and RENAULT, E. (1996) Stochastic volatility. In *Handbook of Statistics 14* (eds G. S. MADDALA and C. R. RAO). Amsterdam: Elsevier, 119–91.
GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 4, 221–37.
HARVEY, A. C. (1998) Long-memory in stochastic volatility. In *Forecasting Volatility in the Financial Market* (eds J. KNIGHT and S. SACHELL). London: Butterworth-Heineman.
HARVEY, A. C. and SHEPHARD, N. (1996) Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429–34.
HASTINGS, W. K. (1970) Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.
HERNÁNDEZ, E. and WEISS, G. (1996) *A First Course on Wavelets*. New York: CRC Press.
HIRANO, K. (1998) Essay on the econometric analysis of panel data. Ph.D. Thesis, Harvard University.
HIRANO, K. (2002) Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica* 70, 781–99.
HOSKING, J. R. M. (1981) Fractional differencing. *Biometrika* 68, 165–76.
HOSKING, J. R. M. (1984) Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research* 20, 1898–908.
JENSEN, M. J. (1999a) Using wavelet to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting* 18, 17–32.
JENSEN, M. J. (1999b) An approximate wavelet MLE of short and long memory parameters. *Studies in Nonlinear Dynamics and Econometrics* 3, 239–353.
JENSEN, M. J. (2000) An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control* 24, 361–87.
DE JONG, P. and SHEPHARD, N. (1995) The simulation smoother for time series models. *Biometrika* 82, 339–50.
KIM, S., SHEPHARD, N. and CHIB, S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–93.

LIU, J. S., WONG, W. H. and KONG, A. (1994) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 81, 27–40.
MACEACHERN, S. M. (1992) Estimating normal means with a conjugate style Dirichlet process prior, Technical report no. 487, Dept. of Statistics, Ohio State University.
MAHIEU, R. J. and SCHOTMAN, P. C. (1998) An empirical application of stochastic volatility models. *Journal of Applied Econometrics* 13, 333–60.
MALLAT, S. (1989) A theory of multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 674–93.
MALLAT, S. (1999) *A Wavelet Tour of Signal Processing* (2nd edn). New York: Academic Press.
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, 1087–92.
MEYER, R. and YU, J. (2000) BUGS for a Bayesian analysis of stochastic volatility models. *Econometric Journal* 3, 198–215.
MCCOY, E. J. and WALDEN, A. T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* 5, 26–56.
MCLEOD, A. I. and HIPEL, K. W. (1978) Preservation of the rescaled adjusted range. *Water Resource Research* 14, 491–508.
NELSON, D. B. (1991) Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica* 59, 347–70.
PERCIVAL, D. B. and WALDEN, A. T. (1998) *Spectral Analysis for Physical Applications*. Cambridge, UK: Cambridge University Press.
PERCIVAL, D. B. and WALDEN, A. T. (2000) *Wavelet Methods for Time Series Analysis*. Cambridge, UK: Cambridge University Press.
RAY, B. K. and TSAY, R. S. (2000) Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics* 18, 254–62.
ROBINSON, P. M. (1995) Log-periodogram regression for time series with long range dependence. *Annals of Statistics* 23, 1048–72.
SCOTT, L. O. (1987) Option pricing when the variance changes randomly: theory, estimation, and an application. *Journal of Financial and Quantitative Analysis* 22, 419–38.
SETHURAMAN, J. and TIWARI, R. C. (1982) Convergence of Dirichlet measures and the interpretation of their parameters. *Statistical Decision Theory and Related Topics* III 2, 305–15.
SHEPHARD, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models* (eds D. R. COX, O. E. BARNDOREF-NIELSEN). London: Chapman and Hall.
SO, M. K. P. (2002) Bayesian analysis of long memory stochastic volatility models. *Sankhya: The Indian Journal of Statistics* 64, 1–10.
TANNER, M. A. and WONG, W. H. (1987) The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, 528–50.
TAYLOR, S. J. (1994) Modeling stochastic volatility: a review and comparative study. *Mathematical Finance* 4, 183–204.
TAYLOR, S. J. (1986) *Modeling Financial Times Series*. New York: Wiley.
TEWFIK, A. H. and KIM, M. (1992) Correlation structure of the discrete wavelet coefficient of fractional Brownian motion. *IEEE Transactions on Information Theory* 38, 904–9.
TIERNEY, L. (1994) Markov chains for exploring posterior distributions. *Annals of Statistics* 21, 1701–62.
TIWARI, R. C., JAMMALAMADAKA, S. R. and CHIB, S. (1988) Bayes prediction density and regression estimation a: semiparametric approach. *Empirical Economics* 13, 209–22.
WHITCHER, B. (1998) Assessing nonstationary time series using wavelets. Ph.D. dissertation, Department of Statistics, University of Washington.
WELCH, B. L. and PEERS, H. W. (1963) On formulae for confidence points based on integrals of weighted likelihoods. *Journal of the Royal Statistical Society, Series* B 25, 318–29.
WEST, M., MÜLLER, P. and ESCOBAR, M. D. (1994) Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D.V. Lindley* (eds A. F. M. SMITH and P. R. FREEMAN). London: Wiley, pp. 363–86.
WRIGHT, J. H. (1999) A new estimator of the fractionally integrated stochastic volatility model. *Economics Letters* 63, 295–303.
WORNELL, G. W. and OPPENHEIM, A. V. (1992) Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing* 40, 611–23.