

Robust Estimation of Nonstationary, Fractionally Integrated, Autoregressive, Stochastic Volatility

Mark J. Jensen

Working Paper 2015-12
November 2015

Abstract: Empirical volatility studies have discovered nonstationary, long-memory dynamics in the volatility of the stock market and foreign exchange rates. This highly persistent, infinite variance—but still mean reverting—behavior is commonly found with nonparametric estimates of the fractional differencing parameter d , for financial volatility. In this paper, a fully parametric Bayesian estimator, robust to nonstationarity, is designed for the fractionally integrated, autoregressive, stochastic volatility (SV-FIAR) model. Joint estimates of the autoregressive and fractional differencing parameters of volatility are found via a Bayesian, Markov chain Monte Carlo (MCMC) sampler. Like Jensen (2004), this MCMC algorithm relies on the wavelet representation of the log-squared return series. Unlike the Fourier transform, where a time series must be a stationary process to have a spectral density function, wavelets can represent both stationary and nonstationary processes. As long as the wavelet has a sufficient number of vanishing moments, this paper's MCMC sampler will be robust to nonstationary volatility and capable of generating the posterior distribution of the autoregressive and long-memory parameters of the SV-FIAR model regardless of the value of d . Using simulated and empirical stock market return data, we find our Bayesian estimator producing reliable point estimates of the autoregressive and fractional differencing parameters with reasonable Bayesian confidence intervals for either stationary or nonstationary SV-FIAR models.

JEL classification: C11, C14, C22

Key words: Bayes, infinite variance, long-memory, Markov chain Monte Carlo, mean-reverting, wavelets

The views expressed here are the author's and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the author's responsibility.

Please address questions regarding content to Mark J. Jensen, Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8019, mark.jensen@atl.frb.org.

Federal Reserve Bank of Atlanta working papers, including revised versions, are available on the Atlanta Fed's website at www.frbatlanta.org. Click "Publications" and then "Working Papers." To receive e-mail notifications about new papers, use frbatlanta.org/forms/subscribe.

1 Introduction

Long-memory, the behavior where the correlation in a times series decays at a slow hyperbolic rate as the interval between observations increases, is a prevalent feature of the volatility of financial returns. Nonparametric estimates of volatility such as realized volatility (Anderson et al., 1999; 2001; 2003) along with different power transformations of absolute returns (Ding et al., 1993; Granger and Ding, 1996) have exhibited long-memory behavior in a wide array of financial asset returns. Fractionally integrated versions of the ARCH and EGARCH model (Baillie et al., 1996; Bollerslev and Mikkelsen, 1996; 1999) and the stochastic volatility model (Bredit et al., 1998 and Harvey, 2011) also find long-memory in the volatility of returns.

Different degrees of long-memory has been found in the estimates of the fractional differencing parameter d for realized volatility where stationary values of d are found taking on values in the range of 0.35 to 0.45 (see Andersen, Bollerslev, Diebold, and Ebens; 2001; Andersen, Bollerslev, Diebold, and Labys; 2001; Andersen, Bollerslev, Diebold and Labys; 2003). Others have found volatility's fractional order of integration to be significantly greater than the mean-reverting and stationary threshold value of $1/2$. Using a FIGARCH model of the S&P 500 composite stock index, Bollerslev and Mikkelsen (1996, 1999) find d to be near 0.6. Hurvich and Ray (2003) find the d in the daily series of the Deutsch Mark-US Dollar exchange rate to equal 0.56 with a fractionally integrated stochastic volatility model. Using a different time period, Harvey (2010) also finds the volatility of the Deutsch Mark-US Dollar exchange rate following a non-stationary, mean-reverting, long-memory process whose variance is infinite by estimating d to be 0.87 with a fractionally integrated stochastic volatility model.

A long-memory process does not have a finite ordered, moving average representation for positive values of d . Instead, its infinite ordered, moving average representation converges when $d < 1/2$, but diverges when $1/2 \leq d < 1$ (see Cheung and Lai, 1993, and Bae et al., 2005). This divergences when d is great than one-half leads to a non-stationary, mean-reverting, long-memory process whose variance is infinite. Such nonstationary values of d present problems for the existing suite of time-varying, conditional, long-memory, variance model estimators. A fractional differencing parameter between one-half and one causes a unit root type volatility model to be unreliable since differencing volatility results in a series that is to close to being non-invertible. Baillie et al. (1996) and Bollerslev and Mikkelsen (1996, 1999) estimate the fractionally integrated ARCH class of models by truncating the long-memory process's infinite moving average average representation. The size of the

approximation error from this truncation grows as the value of d get larger, and, hence, should in general be avoided regardless of d being greater than one-half or not.

Besides volatility not having a variance when d is greater than one-half, the spectral density function of the volatility process also does not exist. Hence, Fourier based semi-parametric estimators of a fractionally integrated stochastic volatility model, for example, Deo and Hurvich (2001), Hurvich and Ray (2003) and Artech (2004), also suffer when d is greater than one-half. Hurvich, Moulines and Soulier (2005) and Frederiksen and Nielsen (2008) overcome a d greater than one-half by computing the pseudo-spectral density function. They prove the frequency domain, local Whittle estimator of d to be a consistent estimator of d when applied to log-squared returns. However, these frequentist, local Whittle estimators are only asymptotically normal when $0 \leq d \leq 3/4$ and have been shown to be heavily biased in simulation studies.

In this paper we are interested in providing the Bayesian joint posterior distribution of the short and long-memory parameters under both stationary and nonstationary, fractionally integrated, autoregressive, stochastic volatility (SV-FIAR). Short memory volatility dynamics are be governed by stationary values of the autoregressive parameter. Stationary, and nonstationary, long-memory volatility behavior is determined by the fractional differencing parameter. We obtain the posterior distribution of the SV-FIAR model's autoregression and long-memory parameters by constructing a highly efficient Markov chain Monte Carlo (MCMC) algorithm in the wavelet domain to produce draws from the parameters joint posterior distribution.

Bayesian methods have become more and more the norm in estimating autoregressive, stochastic volatility models (see the Introduction to Shephard, 2005). This dominance has occurred because the Bayesian approach integrates away the unobservable volatilities to produce likelihood-based estimates of the volatility process' unknown parameters. A Bayesian approach is also appealing because, unlike the above frequentist approaches, which rely on asymptotic results, posterior inference is exact and only conditions on the observed assets returns. These advantages have already been applied by Jensen (2004) in the posterior inference of stationary, long-memory, stochastic volatility, but have not yet been applied to the non-stationary case.

The novelty of our Bayesian approach is its use of the Daubechies (1988) class of wavelets to transform a nonstationary, highly persistent, volatility process into the wavelet domain where its wavelet coefficients are nearly uncorrelated but always with finite variance. While the Fourier series is a basis for stationary processes, wavelets provide a basis for the class of stationary and nonstationary processes. Wavelet analysis is the analysis of change where

the wavelet coefficient measures the amount of information lost at a point in time when a weighted moving average of larger and larger order is applied to the time series; e.g., for a first differenced time series the original time series can be found by adding back the value of the wavelet coefficients to the differenced times series. It is this decorrelating feature along with the wavelet being a basis function for nonstationary processes that makes wavelet analysis ideal for estimating stationary and nonstationary SV-FIAR models.

We apply our Bayesian sampler to stationary and nonstationary SV-FIAR model quasi-return data. We also use the sampler to estimate a SV-FIAR model for daily stock market return data. For the artificially generated SV-FIAR return data our Bayesian estimator produces good point estimates of the autoregressive parameter and of stationary and nonstationary values of the fractional differencing parameter. Empirically, we find strong evidence of non-stationary, mean reverting, long-memory behavior in 30 years worth of volatility dynamics from the Center for Research in Security Prices value-weighted stock portfolio index. Bollerslev and Mikkelesen (1996) also found similar non-stationary, long-memory behavior in the conditional, heteroskedasticity of stock market returns. However, as mentioned above their approach truncates the long-memory process's infinite moving average representation causing their estimate of d to capture only part of the long-memory, persistence of volatility. The approach we propose here does not truncate the moving average representation of long-memory and as such our estimator allows the autoregressive term to model the short-term dynamics of volatility and not be contaminated by the left over persistence not modeled by the truncated approximation.

The contents of the paper are as follows. In Section 2 we introduce the fractionally integrated, autoregressive, stochastic volatility model where volatility is highly persistence, situations where the variance of volatility is infinite but the volatility process is still mean-reverting, and others where volatility nonstationary. Bayesian inference of the stochastic volatility model is explained in Section 3. Section 4 then provides the necessary wavelet theory to construct the paper's sampling algorithm. We recommend either Mallat (1999) or Percival and Walden (2000) to those interested in a more complete description of the statistical theory behind wavelet analysis. Section 5 constructs the Markov chain Monte Carlo simulator using a normal mixture representation the log-squared returns wavelet coefficients unknown distribution. Applications of the sampler are found in Section 6 where the fully parametric Bayesian wavelet estimator is applied to simulated SV-FIAR models and daily compounded stock market return data. Lastly, Section 7 summarizes our findings.

2 SV-FIAR Model

Define the fractionally integrated, autoregressive, stochastic volatility model (SV-FIAR) as

$$y_t = \sigma \exp\{v_t/2\}\xi_t, \quad (1)$$

$$(1 - \phi B)(1 - B)^d v_t = \sigma_\eta \eta_t, \quad t = 1, \dots, T, \quad (2)$$

where at time t the mean corrected return from holding a financial instrument is y_t . The series v_t is the unobservable log-volatility of returns where Eq. (2) shows v_t following a fractionally integrated, first-order autoregressive process. We ignore the possible presence of leverage effects (see Nelson, 1991 and Harvey and Shephard, 1996) by assuming the innovations, ξ_t and η_t , to be uncorrelated, standard normal processes. The parameter σ_η is the standard deviation of log-volatility and σ is the modal instantaneous volatility. To ensure stationarity and invertibility in the autoregressive dynamics of v_t , the absolute value of the autoregressive parameter, ϕ , is assumed to be less than one. Lastly, we denote the parameter vector of the SV-FIAR model as $\beta = (\sigma, \sigma_\eta, d, \phi)'$.

2.1 Stationary behavior

As the product of two independent stochastic processes, one being the Gaussian white noise process, ξ_t , whose variance is known to be one, and the other being $\exp\{v_t/2\}$, all the odd moments of y_t will be zero, whereas its second and fourth order moments depend directly on the corresponding moments of v_t . Because η_t is assumed to be standard normal, $\exp\{v_t/2\}$ is distributed log-normal with a mean of one. If the variance of v_t exists it follows that the unconditional variance, $\text{Var}[y_t] = \sigma^2 \exp\{\sigma_v^2/2\}$, where $\sigma_v^2 \equiv \text{Var}[v_t]$. When the variance of v_t does not exist, y_t continues to be a martingale difference process but its unconditional variance is no longer defined. Hence, stationarity of y_t will depend on the variance of v_t . If the variance of v_t is finite (infinite), then y_t will be a (strictly) stationary process.

Long-memory processes like v_t were independently introduced by Granger and Joyeux (1980) and Hosking (1981).¹ These long-memory processes are defined in terms of the fractional differencing operator, $(1 - B)^d$, whose binomial expansion equals

$$(1 - B)^d = \sum_{j=0}^{\infty} \pi_j(d) B^j, \quad \text{where } \pi_j(d) = \prod_{0 < k \leq j} \frac{k - 1 - d}{k} \sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad \text{as } j \rightarrow \infty,$$

B^j is the lag operator, $B^j x_t \equiv x(t - j)$, $j = 0, 1, 2, \dots$, and $\Gamma(\cdot)$ equals the gamma function.

¹For a review of the work performed in economics with long-memory processes see Baillie (1996).

Inverting the differencing operator results in

$$(1 - B)^{-d} = \sum_{j=0}^{\infty} \psi_j(d) B^j \equiv F(d, 1; 1; B), \quad (3)$$

where $\psi_j(d) = \pi_j(-d) \sim j^{d-1}/\Gamma(d)$, as $j \rightarrow \infty$, and $F(\cdot)$ is the hypergeometric function (see Gradshteyn and Ryzhik, 1994, p. 1066). When $0 < d < 1/2$, the ψ_j s are not absolutely summable, instead, the sequence $\{\psi_j\}$ is square-summable; i.e., $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ (see Beran; 1994).

Combining our assumption of $|\phi| < 1$ with the square summability condition, $0 < d < 1/2$, it directly follows that v_t is a unique, mean zero, covariance stationary process (see Brockwell and Davis; 1993, Theorem 13.2.2). Under these assumptions the latent volatility process, v_t , exhibits the variance

$$\sigma_v^2 = \frac{\sigma_\eta^2}{1 - \phi^2} \frac{\Gamma(1 - 2d)}{\Gamma^2(1 - d)},$$

and the long-memory behavior captured by the slow hyperbolic rate of decay of its autocovariance

$$E[v_t v_{t+s}] \sim |s|^{2d-1} \quad \text{as } s \rightarrow \infty,$$

where $a_s \sim b_s$ indicates $a_s/b_s \rightarrow c$, for some positive constant c , and the exponential rate of increase at frequency zero of its power spectrum

$$\begin{aligned} S_v(\omega) &= \frac{\sigma_\eta^2}{2\pi} \frac{1}{|1 - \phi e^{-i\omega}|^2} |1 - e^{-i\omega}|^{-2d}, \\ &= \frac{\sigma_\eta^2}{2\pi} \frac{1}{|1 - \phi e^{-i\omega}|^2} [4 \sin^2 \omega]^{-d}, \\ &\sim \frac{\sigma_\eta^2}{2\pi} \frac{1}{|1 - \phi|^2} \omega^{-2d} \quad \text{as } \omega \rightarrow 0. \end{aligned} \quad (4)$$

2.2 Strictly stationary behavior

Past research on long-memory stochastic volatility models has primarily focused on the case where v_t is a weakly stationary, long memory process with no autoregressive terms; i.e., $0 < d \leq 1/2$ and $\phi = 0$ (see Deo and Hurvich, 2001; Hurvich et al., 2005; and Jensen, 2004). This papers contributions to the analysis of a (strictly) stationary long-memory, autoregressive, volatility process where $0 \leq d < 1$ ($1/2 \leq d < 1$). When $d \geq 1/2$, the sequence $\{\psi_j\}$ is no longer square-summable, hence, σ_v^2 will be undefined (Granger and

Joyeux, 1980) and neither v_t nor y_t will be weakly stationary. However, it follows from the moving average representation of volatility

$$\begin{aligned}
v_t &= (1 - B)^{-d}(1 - \phi B)^{-1}\eta_t, \\
&= F(d, 1, 1; B)(1 - \phi B)^{-1}\eta_j, \\
&= \sum_{j=0}^i \psi_j \left[\sum_{i=0}^{\infty} \phi^i \eta_{t-i} \right],
\end{aligned} \tag{5}$$

and the definition of the hypergeometric function $F(d, 1, 1; B)$, that when $d < 1$ the effect of the innovations, η_{t-i} , $i = 0, 1, \dots$, dies out over time since the hypergeometric function, $F(d, 1, 1; 1) = 0$, when $d < 1$. In other words, v_t is a mean reverting process even when $1/2 \leq d < 1$. Thus, v_t and y_t are strictly stationary processes when $1/2 \leq d < 1$. On the other hand, the hypergeometric function $F(d, 1, 1; 1)$ diverges when $d \geq 1$, causing past shocks to v_t to persist infinitely into the future.

3 Bayesian analysis

Because y_t is comprised of the two innovations ξ_t and η_t , the likelihood function consists of a mixture distribution over the latent v

$$f(y|\boldsymbol{\beta}) = \int f(y|v, \boldsymbol{\beta})f(v|\boldsymbol{\beta}) dv,$$

where $y = (y_1, \dots, y_T)'$ and $v = (v_1, \dots, v_T)'$. Because v_t enters through the variance of the density of y_t , $f(y|\boldsymbol{\beta})$ has no analytical solution. It follows from the non-analytic nature of the likelihood that the SV-FIAR model has neither a maximum likelihood estimator of $\boldsymbol{\beta}$, nor Bayesian posterior distribution, $\pi(\boldsymbol{\beta}|y) \propto \pi(\boldsymbol{\beta})f(y|\boldsymbol{\beta})$, where $\pi(\boldsymbol{\beta})$ is the parameter vector's prior. Fortunately, this can be overcome with a Bayesian approach where the unknown $\boldsymbol{\beta}$ is augmented with the unobservable volatilities, v , and a Monte Carlo Markov chain (MCMC) sampling algorithm is designed to produce draws from $\pi(\boldsymbol{\beta}, v|y)$ (see Tanner and Wong, 1987; and Tierney, 1994). Drawing $\boldsymbol{\beta}^{(i)}$ and $v^{(i)}$ from $\pi(\boldsymbol{\beta}, v|y)$, where $i = 1, \dots, M$ represent the different draws, posterior inference can be made with the Monte Carlo approximation, $\pi(\boldsymbol{\beta}|y) = \int f(\boldsymbol{\beta}|v, y)f(v|y) dv \approx M^{-1} \sum_{i=1}^M f(\boldsymbol{\beta}|v^{(i)}, y)f(v^{(i)}|y)$ and the law of iterative expectation, $M^{-1} \sum_{i=1}^M \boldsymbol{\beta}^{(i)} \rightarrow E[\boldsymbol{\beta}|y]$, as $M \rightarrow \infty$ (see Gelfand and Smith, 1990).

Designing a MCMC algorithm involves producing draws of $\boldsymbol{\beta}$ and v from the intractable posterior distribution $\pi(\boldsymbol{\beta}, v|y)$. This task can be made tractable by sampling first from the posterior of $\boldsymbol{\beta}$ conditional on a particular value of v , $\pi(\boldsymbol{\beta}|y, v)$, and then drawing v with a multi-state sampler of the posterior distribution of v conditional on the sampled value

of β , in other words, $\pi(v|y, \beta)$. By iteratively sampling from these conditional marginal distributions one produces draws from the joint posterior, $\pi(\beta, v|y)$.²

Sampling from $\pi(v|y, \beta)$ presents a challenge when v is a short-memory process (see Chib et al., 2002), but especially so when v_t is a long-memory process. Faced with the computing and memory costs associated with sampling a long-memory process (see Chan and Palma, 1998), we build on the sampler by Jensen (2004) for the fractionally integrated stochastic volatility model. Our sampler and that of Jensen both utilize the nearly independent behavior of the fractionally differenced process's wavelet coefficients to efficiently sample the latent volatilities. The wavelet coefficients for the SV-FIAR process are independently distributed random variables with a variance equal the value of the power spectrum integrated over limits determined by the support of the wavelet transform.

4 Wavelet representation

Before we construct our multi-state sampler of the latent volatility's wavelet coefficients, we first provide the necessary wavelet theory to understand how and why sampling in the wavelet domain algorithm is robust to nonstationary long-memory.³ 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. Based on the idea of stretching and compressing the basis function wavelets are essentially a band-pass filter whose support at different dilations partitions the frequency space.

A wavelet is defined as any $L^2(\mathfrak{R})$ function, $\psi(t)$, that satisfies the admissibility condition

$$\int_{-\infty}^{\infty} \frac{d\omega}{\omega} |\hat{\psi}(\omega)|^2 < \infty,$$

where $\hat{\psi}$ denotes the Fourier transform of ψ . Since ψ is a square integrable function, the admissibility condition is satisfied only if $\hat{\psi}(0) = 0$, i.e., $\int \psi(t)dt = 0$ and $\hat{\psi}$ has sufficiently fast decay as $|\omega| \rightarrow \infty$.⁴ A useful pedagogical example of a wavelet that facilitate the findings of this paper is the ideal, high-bandpass, wavelet whose frequency response is

$$\hat{\psi}(\omega) = \begin{cases} 1 & \text{if } |\omega| \in (\pi, 2\pi], \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

²For a general introduction to MCMC simulation see Chib (2001) and Chib and Greenberg (1996).

³For a time series based introduction to wavelet analysis see Percival and Walden (2000), but for applications of wavelets to economics and finance see either In and Kim (2013) or the papers in the book by Gallegati and Semmler (2014).

⁴In general, the wavelets used in practice will possess more vanishing moments than required by the admissibility condition.

Let $\psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n)$ represent the dilations and translations of the wavelet ψ , where $j \in \mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ is the scaling parameter and $n \in \mathbf{Z}$ is the translation parameter. It follows from the Fourier properties for dilated and translated functions that $\hat{\psi}_{j,n}(\omega) = 2^{j/2}e^{-i2^j\omega n}\hat{\psi}(2^j\omega)$ so that the frequency response of the dilated ideal, high-bandpass, wavelet in Eq. (6) equals

$$\hat{\psi}_{j,0}(\omega) = \begin{cases} 2^{j/2} & \text{if } |\omega| \in (2^{-j}\pi, 2^{-j+1}\pi], \\ 0 & \text{otherwise.} \end{cases}$$

Because the Fourier transform of the translated wavelet $\psi_{j,n}$ includes the complex exponential term, $e^{-i2^j\omega n}$, $\hat{\psi}_{j,n}$ does not equal $\hat{\psi}_{j,0}$, but their supports are the same.

Daubechies (1988) shows $\{\psi_{j,n}\}_{j,n \in \mathbf{Z}}$, to be a complete orthonormal basis of L^2 such that any $x(t) \in L^2(\mathfrak{R})$ can be represented as

$$x(t) = \sum_{j \in \mathbf{Z}} \sum_{n \in \mathbf{Z}} W_{j,n} \psi_{j,n}(t), \quad (7)$$

where $W_{j,n} = \int x(t)\psi_{j,n}(t) dt$, $j, n \in \mathbf{Z}$, are the wavelet coefficients of $x(t)$.

Drawing on the ideal high-bandpass wavelet, the wavelet representation of $x(t)$ in Eq. (7) essentially partitions the frequency space into the dyadic blocks, $(\pm 2^{-j}\pi, \pm 2^{-j+1}\pi]$. For each positive integer value of the scaling parameter, j , the dyadic frequency block moves down an octave (frequencies twice as small) from those with scale $j - 1$, with half the range of frequencies but with log constant length. Small (large) values of j are associated with dyadic blocks of large (small) support over high (low) frequency octaves.

4.1 Discrete wavelet filter

In theory the wavelet coefficient, $W_{j,n}$, equals the convolution of $x(t)$ with $\psi_{j,n}$, but as an empirical time series $x(t)$ is only observed at a discrete set of observations. Empirically, $W_{j,n}$ is calculated with a two-channel filter bank.⁵ Thus, one never needs an analytical formula for ψ . Instead all the calculations are performed in terms of a filter bank of coefficients.⁶

A two-channel filter bank representation of the wavelet transform consists of the low-pass filter

$$\phi(t) = \sqrt{2} \sum_{k=0}^{L-1} g_l \phi(2t - l), \quad (8)$$

⁵See Strang and Nguyen (1996) for an introduction to wavelets and filter banks.

⁶The wavelet ψ can easily be calculated numerically from the values of these filter bank coefficients (see Daubechies, 1992, p. 205).

where $\{g_l\}_{l=0}^{L-1}$ are non-zero filter coefficients of positive even length L . The matching high-bandpass filter to the wavelet transforms two-channel filter is

$$\psi(t) = \sum_{l=0}^{L-1} h_l \phi(2t - l), \quad (9)$$

where $h_l = (-1)^l g_{L-1-l}$. Daubechies (1988) provides a sufficient set of non-zero values for $\{h_l\}_{l=0}^{L-1}$ where ψ is compactly supported with the smallest possible support, $L - 1$, for a wavelet possessing $L/2$ vanishing moments and whose regularity (number of derivatives and support size) increases linearly with L . Wavelets constructed with the filter coefficients of Daubechies are referred to the Daubechies class of wavelets with order L .

The low-pass filter coefficients, $\{g_l\}$, are equivalent to a moving average filter that smooth the high frequency traits of the series. For the Daubechies class of wavelets these low-pass filter coefficients have the squared-gain function

$$\mathcal{G}(\omega) \equiv \left| \frac{1}{2\pi} \sum_{l=0}^{L-1} e^{-i\omega l} g_l \right|^2 = 2 \cos^L \omega \sum_{l=0}^{L/2-1} \binom{L/2 - l + l}{l} \sin^{2l} \omega. \quad (10)$$

As L increase \mathcal{G} converges to the squared-gain function of a ideal low-pass filter supported on $[-\pi, \pi]$ (see Lai, 1995).

The high-bandpass filter coefficients $\{h_l\}$ constitute a two-stage filter comprised of a first-stage $L/2$ -differencing operator and a second-stage weighted moving average. These two-stages are better understood in terms of the squared gain function of $\{h_l\}$

$$\begin{aligned} \mathcal{H}(\omega) &= (4 \sin^2 \omega)^{L/2} \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \binom{L/2 - l + l}{l} \cos^{2l} \omega, \\ &= \mathcal{D}^{L/2}(\omega) \mathcal{A}_L(\omega), \end{aligned} \quad (11)$$

where $\mathcal{D}^{L/2}(\omega) = (4 \sin^2 \omega)^{L/2}$ is the square modulus of the transfer function for the $L/2$ -order differencing operator, $(1 - B)^{L/2}$, and the squared gain function

$$\mathcal{A}_L(\omega) = \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \binom{L/2 - l + l}{l} \cos^{2l} \omega,$$

is approximately a low-pass filter. As L increases \mathcal{H} approaches the squared gain function of an ideal, high-pass, filter with support on the octave, $(\pi/2, \pi]$.

The function $\phi(\cdot)$ is referred to as the scaling function since it does just that. Like the ideal, high-bandpass, wavelet, one can imagine a ideal, low-bandpass, scaling function whose frequency response is

$$\hat{\phi}(\omega) = \begin{cases} 1 & \text{if } \omega \in [-\pi, \pi], \\ 0 & \text{otherwise.} \end{cases}$$

Note that unlike the wavelet, $\phi(\cdot)$ does not require any vanishing moments. However, we normalize $\phi(\cdot)$ so that $\int \phi(t) dt = 1$, i.e., $\sum_l g_l^2 = 1$.

In a similar manner to the definition of $\psi_{j,n}$, define the dilations and translations of $\phi(\omega)$ to be $\phi_{j,n} = 2^{-j/2}\phi(2^{-j}t - n)$, where $j, n \in \mathbf{Z}$. The filter bank definition of $\psi_{j,n}$ can then be written as:

$$2^{-j/2}\psi(2^{-j}t - n) = 2^{-j/2} \sum_{l=0}^{L-1} h_l \phi(2^{-(j-1)}t - 2n - l).$$

Using this high-bandpass filter definition of $\psi_{j,n}$ it follows that:

$$W_{j,n} = 2^{-j/2} \sum_{l=0}^{L-1} h_l \int x(t) \phi(2^{-(j-1)}t - 2n - l) dt = \sum_{l=0}^{L-1} h_l V_{j-1,2n+l}, \quad (12)$$

where $V_{j,n} = 2^{-j/2} \int x(t) \phi(2^{-j}t - n) dt$ is the scaling coefficient. Thus, computing $W_{j,n}$ in this manner requires knowledge of $\{V_{j-1,n}\}_{n \in \mathbf{Z}}$.

Calculation of the scaling coefficient, $V_{j,n}$, can be performed by writing $\phi_{j,n}$ in terms of the lowpass filter as:

$$2^{-j/2}\phi(2^{-j}t - n) = 2^{-(j-1)/2} \sum_{l=0}^{L-1} g_l \phi(2^{-(j-1)}t - 2n - l).$$

Convoluting $x(t)$ with the above equation we find that $V_{j,n}$ equals:

$$V_{j,n} = 2^{-(j-1)/2} \sum_{l=0}^{L-1} g_l \int x(t) \phi(2^{-(j-1)}t - 2n - l) dt = \sum_{l=0}^{L-1} g_l V_{j-1,2n-l}. \quad (13)$$

Thus, both $V_{j,n}$ and $W_{j,n}$ are calculated recursively from the smallest to the largest scale with the simple multiplication and addition operators of a two-channel filter bank.⁷

As an example of the two-channel filter wavelet transform suppose we observe the time series $x(t)$ at $t = 1, 2, \dots, 2^{\max}$. Let $V_{0,n}$ be the output from applying the low-pass filter to $x(t)$ at the lowest possible scale, i.e., let $V_{0,n} = x(n)$ for $n = 1, \dots, 2^{\max}$.⁸ The value of $W_{j,n}$ and $V_{j,n}$ for $j = 1, 2, \dots, \max$ and $n = 1, 2, \dots, 2^{\max-j}$ are recursively calculated from the $x(t)$ s by applying over and over the filters of Eq. (12) and (13).

This recursive algorithm known as the Fast Wavelet Transform is illustrated in Fig. 1 where $\mathbf{x} = (x(1), \dots, x(2^{\max}))'$, $\mathbf{V}_j = (V_{j,1}, \dots, V_{j,2^{\max-j}})'$, and $\mathbf{W}_j = (W_{j,1}, \dots, W_{j,2^{\max-j}})'$.

⁷Calculation of the wavelet coefficients is even easier than this when the filter coefficients are strategically placed in a sparse matrix. The wavelet transform then becomes a fast $\mathcal{O}(T)$ calculation, which is more efficient than the Fast Fourier Transform's $\mathcal{O}(T \log_2 T)$ calculation.

⁸In general, a function $f \in L^2(\mathfrak{R})$ cannot be sampled as is but must first be smoothed by a low-bandpass filter to eliminate those points where f is possibly undefined. By defining $x(n) = V_{0,n}$, we assume x has been passed through a low-bandpass filter causing the signals values between $x(t)$ and $x(t+1)$ to equal $x(t+1)$ for $t = 1, 2, \dots, 2^{\max} - 1$.

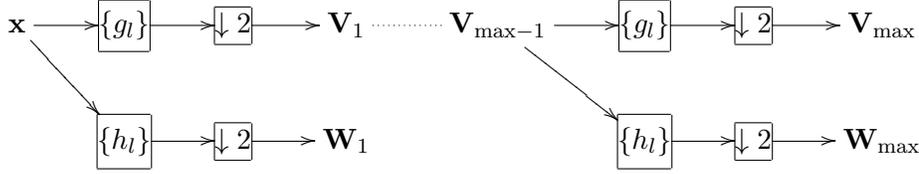


Figure 1: Schematic representation of the Fast Wavelet Transform.

The wavelet coefficients for any value of j represents the information lost when \mathbf{V}_{j-1} is filtered by $\{g_l\}$, i.e., \mathbf{W}_j contains the details or information needed to obtain \mathbf{V}_{j-1} from \mathbf{V}_j . The box $\downarrow 2$ in Fig. 1 represents the decimation of the output from the filter by 2, i.e., discarding the observations with odd time stamps. By their definition the ideal, low, and high-bandpass, filters, $\phi(\cdot)$ and $\psi(\cdot)$, include this decimation. Because the filters coefficients $\{g_l\}$ and $\{h_l\}$ are both applied to \mathbf{V}_j , twice as many observations as the length of \mathbf{V}_j are created. Only half the output from the two-channel filter is needed to completely represent or recover \mathbf{V}_j .

Because of the orthogonality of the filter banks for the Daubechies wavelet, the down arrows in Fig. 1 can be reversed to synthesize \mathbf{x} from its wavelet coefficients, \mathbf{W}_j , $j = 1, \dots, \max$. In the wavelet synthesis, adding the details of \mathbf{W}_j to the smoothed series \mathbf{V}_j provides us with a representation of \mathbf{x} at the next degree of resolution \mathbf{V}_{j-1} with the highest resoluteness being the actual series \mathbf{x} .

Though not as fast the Fast Wavelet Transform in computing the wavelet coefficients, there exists a cleaner definition of $W_{j,n}$ and $V_{j,n}$ in terms of the series $x(t)$ and the filter banks $\{g_l\}$ and $\{h_l\}$. They are

$$W_{j,n} = 2^{-j/2} \sum_{l=0}^{L_j-1} h_{j,l} x \left(2^j(t+1) - 1 - l \right), \quad (14)$$

$$V_{j,n} = 2^{-j/2} \sum_{l=0}^{L_j-1} g_{j,l} x \left(2^j(t+1) - 1 - l \right) \quad (15)$$

where $\{h_{j,l}\}$ and $\{g_{j,l}\}$ are filter coefficients at scale j associated with the Daubechies class of wavelets and $L_j = (2^j - 1)(L - 1) + 1$. Note that for $j = 1$, $h_{1,l} = h_l$ and $g_{1,l} = h_l$. From the definition of $W_{j,n}$ and $V_{j,n}$ in Eq. (12) and (13) and the squared-gain functions of Eq. (10) and (11), it follows that the squared-gain functions of $\{g_{j,l}\}$ and $\{h_{j,l}\}$ equal

$$\mathcal{G}_j(\omega) \equiv \left| \frac{1}{2\pi} \sum_{l=0}^{L_j-1} e^{-i\omega l} g_{j,l} \right|^2 = \prod_{l=0}^{j-1} \mathcal{G}(2^l \omega), \quad (16)$$

$$\mathcal{H}_j(\omega) \equiv \left| \frac{1}{2\pi} \sum_{l=0}^{L_j-1} e^{-i\omega l} h_{j,l} \right|^2 = \mathcal{H}(2^{j-1}\omega) \mathcal{G}_{j-1}(\omega) \quad (17)$$

where $\mathcal{H}_1 = \mathcal{H}$ and $\mathcal{G}_1 = \mathcal{G}$. Since \mathcal{G}_j and \mathcal{H}_j are the product of dilated \mathcal{G} and \mathcal{H} , as L increases \mathcal{G}_j approaches the ideal low-pass filter with the support $[-\pi/2^j, \pi/2^j]$ and \mathcal{H}_j approaches the ideal band-pass filter with support on the intervals $(\pm\pi/2^j, \pm\pi/2^{j-1}]$.

These squared-gain functions also have a nice intuitive flavor to them that follows from the FWT schematic in Fig. 1. Skipping \mathbf{V}_1 and \mathbf{W}_1 , because their squared-gain functions follow directly from (10) and (11), the scaling coefficients \mathbf{V}_2 and wavelet coefficients \mathbf{W}_2 are computed by respectively filtering \mathbf{V}_1 by $\{g_l\}$ and $\{h_l\}$ and then decimating the filtered output by two. Since the filtered output of \mathbf{x} has already been decimated by two in order to obtain \mathbf{V}_1 , applying $\{g_l\}$ and $\{h_l\}$ to \mathbf{V}_1 is equivalent to applying filters with the squared-gain functions $\mathcal{G}(2\omega)$ and $\mathcal{H}(2\omega)$. Since \mathbf{V}_1 is the output from filtering \mathbf{x} with coefficients whose squared-gain function is $\mathcal{G}(\omega)$, the squared-gain function for the filter of \mathbf{x} with output \mathbf{V}_2 is equal to the product of the squared-gain function for the filter \mathbf{x} to \mathbf{V}_1 , and the squared-gain function of the filter \mathbf{V}_1 to \mathbf{V}_2 ; i.e., $\mathcal{G}_2(\omega) = \mathcal{G}(\omega)\mathcal{G}(2\omega)$. Likewise, a filter applied to \mathbf{x} whose output is \mathbf{W}_2 has a squared-gain function equal to $\mathcal{G}(\omega)\mathcal{H}(2\omega)$; i.e., \mathcal{H}_2 . The squared-gain functions generalizes to \mathbf{V}_j and \mathbf{W}_j being the output from filtering \mathbf{V}_{j-1} with $\{g_l\}$ and $\{h_l\}$, respectively.

4.2 Wavelet variance and covariance

Let v_t be the process defined in Eq. (2) but whose fractional differencing parameter d is allowed to take on any positive real values; i.e., v_t can be either (strictly) stationary or nonstationary. Denote by $\mathbf{W}_j^{(v)}$ the $2^{\max}/j$ length wavelet coefficient vector of v_t at scale j constructed with the Daubechies wavelet having $L/2$ vanishing moments. By Percival (1995) Theorem 1, the wavelet coefficients, $W_{1,n}^{(v)}$, where $n = 1, \dots, 2^{\max}/2$, will be a mean zero process with variance

$$\begin{aligned} \text{Var} [W_{1,n}^{(v)}] &= \int \mathcal{H}(\omega) S_v(\omega) d\omega \\ &= \frac{\sigma_\eta^2}{2\pi} \frac{1}{2^{L-1}} \sum_{l=0}^{L/2-1} \left[\binom{L/2-1+l}{l} \right. \\ &\quad \left. \times \int_{-\pi}^{\pi} \frac{1}{|1 - \phi e^{-i\omega}|^2} \cos^{2l} \omega [4 \sin^2 \omega]^{L/2-d} d\omega \right]. \end{aligned} \quad (18)$$

From the findings of Section 2.2 we know the variance of v_t does not exist when $d \geq 1/2$. However, when $d \leq L/2$ the integrand $[4 \sin^2 \omega]^{L/2-d}$ is bounded by one so the variance of $W_{1,n}^{(v)}$ will be finite. Because a wavelet requires $L \geq 2$, the Daubechies class of wavelets

transforms v_t into a stationary process of wavelet coefficients regardless of v_t being stationary ($d < 1/2$) or strictly stationary ($1/2 \leq d < 1$). Furthermore, when $d \geq 1$, in other words, when v_t is a nonstationary explosive process, $W_{1,n}^{(v)}$ will be stationary with finite variance when the wavelet transform is performed with any L -ordered, Daubechies, wavelet where the number of vanishing moments, $L/2$, is greater than d . This stationarity result generalizes to wavelet coefficients having scale parameters $j > 1$ where

$$\text{Var} [W_{j,n}^{(v)}] = \int \mathcal{H}_j(\omega) S_v(\omega) d\omega.$$

See Percival (1995, Section 6).

From Eq. (18) the variance of $W_{1,n}$ is independent of the value n ; i.e., all the wavelet coefficients of equal scale j have the same variance. Wavelet coefficients of the same scale j are thus homoskedastic but hetroskedastic over different scales. Empirical work with $W_{j,n}$ thus requires computing this variance at each scale, j , using a analytic formula similar to Eq. (18). Given the summation operator this would involve numerically evaluating the integral a number of times for different values of l . A very computationally costly calculation.

An alternative method to computing the variance of $W_{j,n}^{(v)}$ is to recall that \mathcal{H}_j is approximately equal to the squared-gain function of an ideal bandpass filter supported on the octave $(\pi/2^j, \pi/2^{j-1}]$. Employing this approximation in the calculation of the wavelet variances leads to numerically evaluating:

$$\text{Var}[W_{j,n}] \approx 2^{j+1} \int_{2^{-j}\pi}^{2^{-j+1}\pi} S_v(\omega) d\omega. \quad (19)$$

This approximation becomes more exact as L increases, or in other words, the number of vanishing moments $L/2$ increases. However, simulation studies have shown that even for small values of L this approximation is very good (Percival and Walden, 2000, Ch. 8).

Drawing on the ideal bandpass nature of the wavelet, McCoy and Walden (1996) and Jensen (1999, 2000) show that, like the Fourier function, the wavelet is a near diagonalizing operator to a stationary ARFIMA processes covariance matrix. If the number of vanishing moments of the wavelet exceed the differencing parameter d the wavelet coefficient's covariance between and across scales is equal to zero. As a result we can treat the wavelet coefficients $W_{j,n}^{(v)}$ as being a nearly independent stationary process with mean zero and variance approximately equal to Eq. (19).

5 Bayesian analysis with wavelets

The SV-FIAR model can be converted into a infinite dimensional, state-space model by squaring the returns, y_t , and applying the logarithmic transformation, to obtain the SV-FIAR model's linear offset representation:

$$y_t^* = \log \sigma^2 + v_t + z_t \quad (20)$$

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 a offset representation with $c = 0.0005$.

We project the SV-FIAR linear offset series, y_t^* , into the time-scale space of the wavelet domain and arrive at the linear relationship between the wavelet coefficients of y^* and v :

$$W_{j,n}^{(y^*)} = W_{j,n}^{(\log \sigma^2)} + W_{j,n}^{(v)} + W_{j,n}^{(z)}, \quad (21)$$

where $j = 1, \dots, \max \equiv [\log_2 T]$, with $[\cdot]$ being the integer portion of the argument, $n = 1, \dots, T/2^j$, and $W_{j,n}^{(y^*)}$, $W_{j,n}^{(\log \sigma^2)}$, $W_{j,n}^{(v)}$ and $W_{j,n}^{(z)}$ are the wavelet coefficients of y^* , $\log \sigma^2$, v and z , respectively. Note that when the number of observations T is a power of two the number of dilated and translated wavelet coefficients equals $T - 1$. For convenience we assume that T is always a power of 2, however, when the data set is not a power of 2 in length we follow the practice of Ogden (1997) and pad the actual data up to a power of 2 by appending in reverse the observed data.

Begin constant unknown parameter, $\log \sigma^2$ wavelet coefficients are all zero.⁹ Hence, we drop $W_{j,n}^{(\log \sigma^2)}$ from Eq. (21) and write $W_{j,n}^{(y^*)}$ as

$$W_{j,n}^{(y^*)} = W_{j,n}^{(v)} + W_{j,n}^{(z)}. \quad (22)$$

As a 'near' diagonalizing operator the wavelet transformed latent volatilities are normally distributed¹⁰

$$\mathbf{W}^{(v)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{W^{(v)}}), \quad (23)$$

⁹Information on the unknown σ is contained in the scaling coefficient $V_{\max,1}^{(\log \sigma^2)}$, so one could estimate σ by synthesizing the scaling coefficient. Because our primary interest is in estimating values of d that cause volatility to be nonstationary we leave the estimation of σ for future research.

¹⁰See Jensen (2000) for the normality results.

where $\mathbf{W}^{(v)} = (\mathbf{W}_1^{(v)'}, \dots, \mathbf{W}_{\max}^{(v)'})'$. From the results of Section 4.2 on the wavelet variances and covariances of a fractionally integrated, autoregressive process, the wavelet coefficients covariance matrix is approximately

$$\boldsymbol{\Sigma}_{\mathbf{W}^{(v)}} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{\max}^2 \end{bmatrix},$$

with

$$\sigma_j^2 = \frac{2^{j+1} \sigma_\eta^2}{2\pi} \int_{2^{-j}\pi}^{2^{-j+1}\pi} \frac{1}{|1 - \phi e^{-i\omega}|^2} |1 - e^{-i\omega}|^{-2d} d\omega, \quad j = 1, \dots, \max, \quad (24)$$

and \mathbf{I}_j is the $T/2^j \times T/2^j$ identity matrix.

By projecting y^* into the time-scale space of the wavelet basis, Section 3 objective of drawings from $\pi(\boldsymbol{\beta}, v|y)$ changes to designing a MCMC sampler that draws $\boldsymbol{\beta} = (\phi, d, \sigma_\eta^2)'$ and $\mathbf{W}^{(v)}$ from the SV-FIAR wavelet domain posterior distribution, $\pi(\boldsymbol{\beta}, \mathbf{W}^{(v)} | \mathbf{W}^{(y^*)})$, given the observed log-squared returns entire vector of wavelet coefficients, $\mathbf{W}^{(y^*)}$. The near independent, multivariate, Gaussian distribution of $\mathbf{W}^{(v)}$ affords us an easy and efficient way of sampling the latent volatilities. Instead of inefficiently drawing the highly correlated v , the volatilities wavelet coefficient vector, $\mathbf{W}^{(v)}$, can be quickly and efficiently sampled from the posterior

$$\pi(W_{j,n}^{(v)} | W^{(y^*)}, \boldsymbol{\beta}) \propto \pi(W_{j,n}^{(v)} | \boldsymbol{\beta}) f(W^{(y^*)} | W_{j,n}^{(v)}, \boldsymbol{\beta}),$$

where $f(W^{(y^*)} | W_{j,n}^{(v)}, \boldsymbol{\beta})$ is the likelihood function and the conditional prior $\pi(W_{j,n}^{(v)} | \boldsymbol{\beta})$ is defined by Eq. (23). We will detail the first and second moments of this conditional normal posterior distribution in Section 5.3, but for now it should be clear that the independent wavelet coefficients, $W_{j,n}^{(v)}$, enables us to sidestep the difficult task of sampling the highly correlated v_t .

5.1 Sampling distribution

The conditional distributions, $\boldsymbol{\beta} | W^{(y^*)}, W^{(v)}$ and $W^{(v)} | W^{(y^*)}, \boldsymbol{\beta}$ are intractable since the sampling distribution of $W_{j,k}^{(y^*)}$ is not normal. Unlike the time-domain where the z_t in Eq. (20) is known to be distributed log-chi-square, the distribution of $W_{j,k}^{(z)}$ is not known. Thus, we are unable to use the method of Kim et al. (1998) to set the parameters of a mixture of normal distributions to match the first four moments of $\pi(W_{j,k}^{(z)})$. Instead, we

choose to use a Bayesian semiparametric estimator with a Dirichlet process prior to estimate and fix the approximate mixture distribution representation of $\pi(W^{(z)})$.¹¹

A sample of z with $T = 4096$ observations is generated by applying the log-square transformation to a random draw from the standard normal distribution and adding 1.2704 to the simulate series. The wavelet transform is then applied to z to produce a realization of $\mathbf{W}^{(z)}$. A nonparametric Bayesian estimator with a Dirichlet process prior applied to $\mathbf{W}^{(z)}$ results in a second-order mixture of normal distributions where the mixture probabilities are $\pi = (0.798, 0.202)'$ and the means and variances of the two mixture clusters are $\mu = (0.269, -0.994)'$ and $\nu = (1.732, 3.245)'$, respectively. Thus, the date generating process for $W^{(z)}$ is approximated by the normal mixture distribution

$$W_{j,n}^{(z)} \sim 0.798 \mathcal{N}(0.269, 1.732) + 0.202 \mathcal{N}(-0.994, 3.245).$$

Under this mixture representation of $W_{j,k}^{(z)}$ distribution, the likelihood for $\mathbf{W}^{(y^*)}$ equals

$$f\left(\mathbf{W}^{(y^*)} \mid \mathbf{W}^{(v)}, s\right) = \prod_{j=1}^{\max T/2^j} \prod_{n=1} f_N\left(W_{j,n}^{(y^*)} \mid W_{j,n}^{(v)} + \mu_{s_{j,n}}, \sigma_j^2 + \nu_{s_{j,n}}\right), \quad (25)$$

where $s_{j,n} = 1, 2$, such that $\pi_1 \equiv \text{Prob}(s_{j,n} = 1) = 0.798$ and $\pi_2 \equiv \text{Prob}(s_{j,n} = 2) = 0.202$, is the mixture cluster assignment variable and the assignment vector is $s = \{s_{j,n}\}$, where $j = 1, \dots, \max$ and $n = 1, \dots, 2^{\max}/j$. As the assignment variable of the j -scale, n -translated, wavelet coefficient of y^* to a particular mixture cluster, the $s_{j,n}$ are unknown binomial distributed random variables whose prior is the above discrete mixture probabilities.

5.2 Prior

To complete our Bayesian sampler we still need to specify the form of the prior $\pi(\beta)$. We first assume that all the unknown parameters are mutually independent; i.e.,

$$\pi(\beta) = \pi(\sigma_\eta^2)\pi(d)\pi(\phi).$$

To ensure the value of d can produce both stationary and nonstationary, long-memory behavior in volatility we choose an non-informative, uniform prior for $\pi(d)$ with the support $(0, 2)$. For the autoregressive parameter ϕ we also choose a non-informative prior that is uniform over the region where autoregressive processes are stationary, $|\phi| < 1$. Lastly, we assume our initial knowledge concerning σ_η^2 is accurately reflected by the diffuse, inverse gamma distribution

$$\pi(\sigma_d^2 \mid v_0, \delta_0) \propto (\sigma_d^2)^{-(v_0+2)/2} \exp\left\{-\frac{\delta_0}{2\sigma_d^2}\right\}, \quad (26)$$

¹¹See Hirano (2002), Jensen (2004) and Jensen and Maheu (2010) for other econometric problems where the Bayesian nonparametric Dirichlet process prior has been applied.

with hyperparameters $\delta_0 = 2$ and $\nu_0 = 0.02$.

5.3 Sampler

To sample from the desired joint posterior distribution, $\beta, \mathbf{W}^{(v)}, s \mid \mathbf{W}^{(y^*)}$, we design a MCMC sampler that generalizes the Metropolis-Hastings (MH) algorithm found in Jensen (2004). Here we compute the latent wavelet coefficient's variances by numerically evaluating the integral found in Eq. (24). Our hybrid MCMC algorithm is as follows

1. Initialize $s, \beta, \mathbf{W}^{(v)}$.
2. Jointly sample β and $\mathbf{W}^{(v)}$ from $\beta, \mathbf{W}^{(v)} \mid \mathbf{W}^{(y^*)}, s$ by drawing,
 - (a) β from $\beta \mid \mathbf{W}^{(y^*)}, s$, and then drawing,
 - (b) $\mathbf{W}^{(v)}$ from $\mathbf{W}^{(v)} \mid \mathbf{W}^{(y^*)}, \beta, s$.
3. Sample s from $s \mid \mathbf{W}^{(y^*)}, \mathbf{W}^{(v)}$.
4. Repeat Steps 2 and 3.

The MCMC algorithm begins in Step 1 with the initialization of the mixture distribution's assignment vector, s , parameter vector, β , and latent volatilities wavelet coefficients, $\mathbf{W}^{(v)}$. Step 2 is the method of composition approach to jointly sampling, β and $\mathbf{W}^{(v)}$. In Step 2(a), β is sampled from $\beta \mid \mathbf{W}^{(y^*)}, s$, where $\mathbf{W}^{(v)}$ has been marginalized out, and in Step 2(b), $\mathbf{W}^{(v)}$ is drawn from $\mathbf{W}^{(v)} \mid \mathbf{W}^{(y^*)}, s$. This reduced blocking scheme of drawing $\beta, \mathbf{W}^{(v)} \mid \mathbf{W}^{(y^*)}, s$ eliminates the functional dependency that exists between $\mathbf{W}^{(v)}$ and its parameters, β . In contrast, sampling from $\beta \mid \mathbf{W}^{(y^*)}, \mathbf{W}^{(v)}, s$, and $\mathbf{W}^{(v)} \mid \mathbf{W}^{(y^*)}, \beta, s$, produces draws of $\mathbf{W}^{(v)}$ and β that are strongly correlated. Lastly, Step 3 in the above sampler, involves independently sampling the $s_{j,n}$, $j = 1, \dots, \max$, $n = 1, \dots, 2^{\max-j}$, from the set $\{1, 2\}$ where the probability of selecting the i th element equals

$$P\left(s_{j,n} = i \mid \mathbf{W}_{j,n}^{(y^*)}, \mathbf{W}_{j,n}^{(v)}\right) \propto \pi_i f_{\mathcal{N}}\left(\mathbf{W}_{j,n}^{(y^*)} \mid \mathbf{W}_{j,n}^{(v)} + \mu_i, \nu_i\right), \quad i = 1, 2,$$

where $\pi_1 = 0.789$ and $\pi_2 = 0.202$.

The conditional distribution, $\beta \mid \mathbf{W}^{(y^*)}, s$, in Step 2(a) is proportional to

$$\pi(\beta \mid \mathbf{W}^{(y^*)}, s) \propto \pi(\beta) f(\mathbf{W}^{(y^*)} \mid s, \beta), \quad (27)$$

where from the likelihood function in Eq. (25)

$$f(\mathbf{W}^{(y^*)} \mid s, \beta) = \prod_{j=1}^J \prod_{n=1}^{T/2^j} \frac{1}{\sqrt{2\pi(\nu_{s_{j,n}}^2 + \sigma_j^2)}} \exp\left\{-\frac{(W_{j,n}^{(y^*)} - \mu_{s_{j,n}})^2}{2(\nu_{s_{j,n}}^2 + \sigma_j^2)}\right\}, \quad (28)$$

with σ_j^2 being defined in Eq. (24).

Eq. (27) is a nonstandard distribution and hence, draws from $\beta|\mathbf{W}^{(y^*)}, s$ cannot be made from a standard distribution. We choose to employ the tailored Metropolis-Hasting (MH) algorithm to sample this intractable distribution.¹² With the tailored MH sampler we make a candidate draw of $\beta|\mathbf{W}^{(y^*)}, s$ by sampling from the multivariate Student- 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- t mean and covariance equal to

$$\begin{aligned} m &= \arg \max_{\beta} \log \pi(\beta) f(W^{(y^*)}|s, v, \beta), \\ V &= [-\partial^2 \log \pi(\beta) f(W^{(y^*)}|s, v, \beta) / (\partial\beta \partial\beta') \Big|_{\beta=m}]^{-1}, \end{aligned}$$

whose values are obtained with the quasi-Newton maximization algorithm developed by Broyden, Fletcher, Goldfarb, Shanno initialized at the value of β from the MCMC samplers previous sweep. The degrees of freedom is set to $\eta = 10$.

Denote the candidate draw from the Student- t distribution as β' . The draw, β' , is accepted as a realization from $\beta|\mathbf{W}^{(y^*)}, s$ with Metropolis-Hastings probability

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

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

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

$$\pi(W_{j,n}^{(v)} | W_{j,n}^{(y^*)}, \beta, s_{j,n}) \propto \pi(W_{j,n}^{(v)} | \beta) f(W_{j,n}^{(y^*)} | W_{j,n}^{(v)}, \beta, s_{n,j}),$$

where $\pi(W_{j,n}^{(v)} | \beta) \sim \mathcal{N}(0, \sigma_j^2)$ and the likelihood function is

$$\begin{aligned} f(W_{j,n}^{(y^*)} | W_{j,n}^{(v)}, \beta, s) &= \frac{1}{\sqrt{2\pi(\sigma_j^2 + \nu_{s_{j,n}}^2)}} \\ &\times \exp \left\{ -\frac{(W_{j,n}^{(y^*)} - W_{j,n}^{(v)} - \mu_{s_{j,n}})^2}{2(\sigma_j^2 + \nu_{s_{j,n}}^2)} \right\}, \end{aligned}$$

¹²See Chib and Greenberg 1995, 1998 for the theory behind MH samplers.

for $j = 1, \dots, \max$, and $j = 1, \dots, 2^{\max-j}$. After completing the square the conditional latent wavelet coefficients is found to be independently distributed

$$W_{j,n}^{(v)} | W_{n,j}^{(y^*)}, \beta, s_{n,j} \sim \mathcal{N}(\bar{W}_{j,n}, v_{j,n}^2),$$

where

$$\bar{W}_{j,n} = \left(\frac{1/\sigma_{s_{j,n}}^2}{1/\sigma_{s_{j,n}}^2 + 1/\sigma_j^2} \right) \bar{W}_{j,n}^{(y^*)}, \quad v_{j,n}^2 = \frac{1}{1/\sigma_{s_{j,n}}^2 + 1/\sigma_j^2},$$

and $\bar{W}_{j,n}^{(y^*)} = W_{j,n}^{(y^*)} - \mu_{s_{j,n}}$.

6 Applications

In this section we report the results from our MCMC algorithm when applying it to simulated and empirical daily stock return data. Daubechies (1992, p.198) least asymmetric wavelet with $L = 8$ nonzero filter coefficients (four vanishing moments) is used in every instance. A number of issues need to be taken into consideration when choosing the particular family of wavelets and its number of vanishing moments; issues like boundary effects, possible statistical artifacts caused by a wavelet with a large number of vanishing moments, and how well the wavelet approximates an ideal bandpass filter. We choose the Daubechies least asymmetric class of wavelets with four vanishing moments because it favorably addresses each of these items while possessing the extraordinary theoretical property of diagonalizing a long-memory process's covariance matrix (Whitcher et al.; 2000 and Percival and Walden; 2000, pp. 346-349).

6.1 Quasi-returns

We generate and estimate a collection of SV-FIAR models having $T = 4096$ observations and consisting of twelve different combinations of the autoregressive and differencing parameter values; $\phi = 0, 0.5, 0.75, 0.99$ and $d = 0, 0.45, 0.6, 1.0$.¹³ These particular values of ϕ and d generate a wide variety of time series behavior in the volatility process, v_t , including weak to strong persistent autoregressive dynamics, stationary and strict stationary, long-memory, and explosive, nonstationary unit root behavior. Each of the twelve simulated return series is generated using the instantaneous volatility value of $\sigma^2 = 1$ and the volatility of volatility $\sigma_\eta^2 = 0.01$. In robustness tests, we found that the estimates of d and ϕ were unaffected by larger values for σ and σ_η but we choose not to report them here.

¹³Hosking (1984) algorithm is used to generate the latent volatility processes.

Table 1 lists the parameter estimates and their summary statistics for the twelve different data generating SV-FIAR models. Each simulated SV-FIAR model result is separated in the table by horizontal lines. To encourage convergence of the MCMC algorithm to the target posterior density, for each simulated series, only the last 5,000 draws of β and $\mathbf{W}^{(v)}$ out of 6,000 MCMC sweeps are used to make inference concerning the parameter values. In the second through sixth column of Table 1 we list the model’s true parameter values, σ_η^2 , d and ϕ_1 , their posterior mean which equal the sample mean of the MCMC draws, their numerical standard error (NSE), the sampler’s inefficiency measure (Ineff.) for the particular parameter, and the 5 and 95 percentiles (in parenthesis) of the parameter’s posterior distribution. Finishing out the table in the seventh column is the tailored Metropolis-Hasting rejection rate, $1 - \alpha(\beta, \beta')$, for the candidate draw from $\beta \sim \beta|W^{(y^*)}, s$.

The inefficiency measure indicates how slow our MCMC sampler converges to the target posterior density by measuring the level of dependency between the drawn β s. The higher the inefficiency the greater the number of draws of β needed to reach a particular level of Monte Carlo accuracy. The inefficiency measure for a parameter is:

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

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 ($N = 5,000$), and J 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 $J = 100$ is adequate.).

The measure of inefficiency quantifies the loss associated with using correlated draws to make inference concerning the parameter as opposed to truly independent draws. This inefficiency measure for each unknown is also used to compute the numerical standard error (NSE) of the parameter’s posterior mean by taking the square root of the product between the parameter’s inefficiency and the sample variance of its draws (Geweke, 1992).

The parameter inefficiency measures and MH rejection rates in Table 1 are indicative of a MCMC sampler that is producing weakly correlated draws from over the entire support of the posterior distribution. In the simulations the efficiency measures are no bigger than thirty and in most cases less than ten. Small inefficiency measures support our decision to jointly sample the latent volatilities wavelet coefficients and the parameters with the two step composition approach of first sampling from $\beta|W^{(y^*)}, s$ with the tailored MH algorithm followed by individually drawing the latent volatilities wavelet coefficients.

$$y_t = \exp\{v_t/2\}\xi_t, (1 - \phi B)(1 - B)^d v_t = \sqrt{0.01} \eta_t, t = 1, \dots, 4096$$

	Truth	Mean	NSE	Ineff.	90% Conf.	$1 - \alpha(\beta, \beta')$
d	0	0.1937	0.0317	6.180	(0.0630, 0.3372)	0.375
ϕ	0.99	0.9715	0.0007	7.593	(0.9489, 0.9870)	
d	1.0	1.0156	0.0187	7.185	(0.9140, 1.1232)	0.453
ϕ	0	0.3080	0.2847	9.469	(-0.0510, 0.6187)	
d	0.45	0.2450	0.0987	10.367	(0.0384, 0.4094)	0.427
ϕ	0.75	0.8700	0.0291	14.486	(0.7812, 0.9479)	
d	0.45	0.3631	0.1482	17.405	(0.1538, 0.5174)	0.516
ϕ	0.5	0.5873	0.3581	18.247	(0.3018, 0.8630)	
d	0.45	0.4082	0.0157	2.592	(0.2368, 0.5487)	0.284
ϕ	0	-0.1277	0.1667	2.734	(-0.5802, 0.3947)	
d	0.6	0.4925	0.0441	6.677	(0.3255, 0.6476)	0.299
ϕ	0.75	0.8387	0.0183	9.588	(0.7330, 0.9141)	
d	0.6	0.4990	0.1607	19.875	(0.2862, 0.6424)	0.538
ϕ	0.5	0.7242	0.2258	26.533	(0.5649, 0.8919)	
d	0.6	0.6211	0.0231	5.766	(0.4949, 0.7410)	0.371
ϕ	0	0.0535	0.3688	7.364	(-0.3726, 0.4621)	

Table 1: Bayesian estimates (6000 draws with the first 1000 draws discarded) of the autoregressive and fractional differencing parameter for a simulated SV-FIAR model where NSE is the numerical standard error, Ineff. is the inefficiency factor, 90% Conf. denotes the 5 and 95 percentiles of the posterior distribution, and $1 - \alpha(\beta, \beta')$ is the MH rejection rate. Note that some of the results are from misspecified models when the true value of either ϕ or d is zero.

The tailored MH portion of the sampler mixes well with the conditional sampler of the parameter vector accepting nearly half of the candidate draws of d and ϕ . Acceptance rates of this size for a distribution with the dimension of $\beta|\mathbf{W}^{(y^*)}$, s are indicative of an accurate Metropolis-Hastings algorithm that has converged and is drawing from the desired target distribution. It also shows that the sampler is not getting stuck in any particular region of the distribution but is spanning the distributions entire domain.

From the parameter estimates reported in Table 1 our Bayesian estimator is capable of distinguishing between a unit root ($d = 1$) in volatility and strongly persistent autoregressive behavior ($\phi = 0.99$). When the simulated returns are generated from a SV-FIAR model with $d = 0$ and $\phi = 0.99$, the sampler correctly assigns the autoregressive persistence to ϕ and not d by finding ϕ to equal 0.9715.¹⁴ At the other end of the persistence spectrum, when the return series is generated with a volatility process where $d = 1$ and $\phi = 0$, our estimate of the differencing parameter captures the nonstationary unit root behavior of volatility by estimating $d = 1.0156$ and $\phi = 0.3080$.

The ability to distinguish between, and assign autoregressive and long memory persistence to their rightful parameter, also holds when ϕ and d move away from the extreme cases of $d = 1$, $\phi = 0$, and $d = 0$, $\phi = 0.99$. However, as the level of persistence assigned to ϕ increases a portion of the long-memory dynamics gets assigned by the estimator to ϕ and not to d . This reallocation of persistence away from d to ϕ causes d to be under estimated. Examples of this are seen in the quasi-returns where $\phi = 0.75$. When volatility is generated with $\phi = 0.75$ and $d = 0.45$, the fractional differencing parameter estimate is $d = 0.245$. When the true value of d is increased to 0.6, our estimate of d also increases but only to 0.4925. This is still less than the true value but noticeably less than when the true fractional differencing parameter was $d = 0.45$.

Under estimating d also occurs, but to a smaller degree, when volatility is simulated with $\phi = 0.5$. Estimates of d when volatility is generated with $\phi = 0.5$ and $d = 0.45, 0.6$ are respectively, 0.3631 and 0.499. The only simulated series where the fractional differencing parameter is over estimated occurs when volatility is void of an autoregressive term ($\phi = 0$). Even though misspecified with a non-zero autoregressive parameter, the estimate of d and ϕ are still very close to their true values. Hence, the simulation results suggest our estimator is capturing some of the strongly persistent, low frequency, long-memory dynamics of volatility in its estimate of ϕ . The residual low frequency energy not estimated by ϕ is then captured in the estimate of d .

¹⁴Since the prior for d is restricted to the positive real line, its true value is outside the support of its posterior distribution. Hence, a prior that contains the origin would most likely produce a smaller estimate of d .

Our wavelet based estimator of the fractional differencing parameter of latent volatility is also robust to values of d that lead to volatility being nonstationary; i.e., when d is greater than $1/2$. For each of the quasi SV-ARFI returns simulated with $d = 0.6, 1.0$ the 90% posterior probability interval of d contains the true value of d . In addition, when volatility is simulated with a nonstationary, long-memory model with no short-memory autoregressive term the SV-FIAR model’s posterior mean of d is within 0.02 of its true value. From these findings one can set aside their worries about over differencing the log-squared returns when estimating a nonstationary, stochastic volatility process with our time-scale wavelet space estimator.

The posterior draws $\{d^{(i)}\}$ and $\{\phi^{(i)}\}$ also provide good inference concerning the autoregressive and long-memory dynamics of volatility. In only three out the twelve simulated SV-FIAR return series does the 90% Bayesian probability interval fail to contain the true parameter value. Two of these are the probability intervals of ϕ when the true autoregressive parameter is 0.99 and 0.5. The third is the probability interval for d when the fractional differencing parameter is 0.45 and the autoregressive parameter was 0.75. Even in these three instances the 90% probability interval for the other parameter contains that parameters actual value. So the posterior uncertainty around the parameter estimate can, in general, be trusted to separate out autoregressive dynamics from long-memory.

6.2 Market returns

In this section we report the results from applying our nonstationary robust, Bayesian estimator of the SV-FIAR model to the continuously compounded daily returns of the Center for Research in Security Prices (CRSP) value weighted portfolio index. The returns are corrected for the effects of stock splits and dividends and consist of 8,192 observations from July 22, 1974 to December 29, 2006.¹⁵

In Table 2 we report the summary statistics of the MCMC draws for the SV-FI model, where the autoregressive parameter ϕ is set equal to zero, and the SV-FIAR model. A total of 6000 sweeps of the MCMC algorithm are made of which the final 5000 draws are used in our posterior analysis of d and ϕ . The MH-rejection rates in Table 2 suggest the β sampler is mixing in a similar manner to the quasi-return cases. The rejection rate is 21% for the SV-FI model and 31% when ϕ is estimated. By being small the inefficiency measures for d and ϕ are again giving us confidence that our sampler is working efficiently and producing

¹⁵As mentioned before the Fast Wavelet transform algorithm requires the sample size to be an integer power of two. For data sets smaller than the next integer power of 2 we increase its length by reflecting the series last observations onto the data set.

Model	β	Mean	NVAR	Ineff.	90% Conf.	$1 - \alpha(\beta, \beta')$
SV-FI	d	0.5896	0.0018	1.8660	(0.5291, 0.6505)	0.2053
SV-FIAR	d	0.6565	0.0050	3.2559	(0.5799, 0.7336)	0.3133
	ϕ	-0.5634	0.0907	4.5891	(-0.7945, -0.2455)	

Table 2: SV-FIAR model estimates for compounded daily returns of the Center for Stock Return Prices value weighted portfolio over the period from July 22, 1974 to December 29, 2006 ($T = 8, 192$).

draws from the entire posterior distribution.

In Fig. 2 the posterior draws of d for the SV-FI model are plotted, along with the histogram, and the first 20 lags of the draws correlogram. Fig. 3 graphs the posterior draws, histogram and lagged autocorrelations for the d and ϕ from the SV-FIAR model. Since the efficiency factors are less than 5, the decay in the correlograms of Fig. 2 and 3 show very little correlations between the parameter draws. This give us comfort that our sampler has converged to the underlying posterior distribution and statistical inference on the unknown value of d and ϕ is possible with these draws.

From the posterior draws of the SV-FI model’s fractional differencing parameter the density of the posterior distribution of d is centered around 0.5896 with a 90% Bayesian probability interval of (0.5291, 0.6505). Such values for the fractional differencing parameter suggests volatility is more persist than had previously been thought, so much so that volatility is a mean-reverting, long-memory process but whose variance is infinite.

When the autoregressive coefficient is included in the model, the value of d increases. This suggests volatility of the stock market is even more persistent and nonstationary than inferred by the SV-FI model. Past short-memory, auto-regressive, stochastic volatility models of the equity market have generally found ϕ to be near one. Hence, one might expect that by including the autoregressive parameter ϕ in the SV-FIAR model it would take away some of the persistence associated with the fractionally differencing operator. However, the opposite occurs. In the SV-ARFI model d increases to 0.66 and the 90% probability interval of d lies over the larger values of (0.59, 0.73). The posterior mean of the autoregressive ϕ , equals -0.56 and the 90% probability interval of $(-0.79, -0.25)$ infers a short-memory dynamics opposite to that of persistence. The histograms of ϕ in Fig. 3 suggest its posterior distribution is not even supported over positive values and hence,

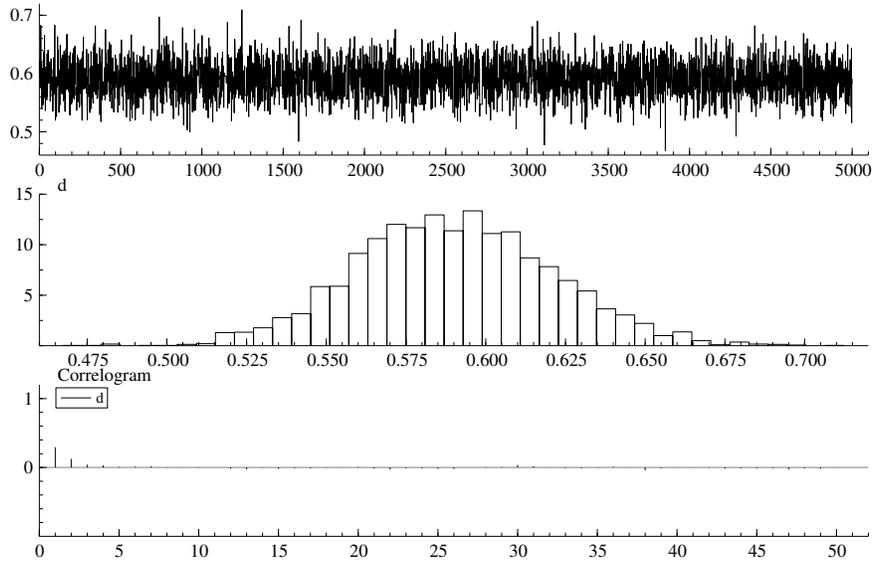


Figure 2: The 5000 MCMC draws of d , their histogram and correlogram for the SV-FI model of daily compounded returns of the Center for Stock Return Prices value weighted portfolio over the period from July 22, 1974 to December 29, 2006 ($T = 8,192$).

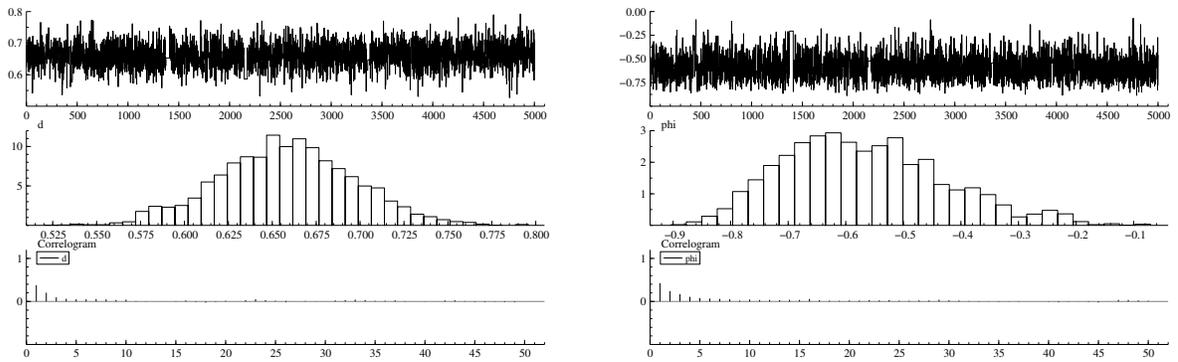


Figure 3: The 5000 MCMC draws of d and ϕ_1 , their histograms and correlograms for the SV-FIAR model of daily compounded returns of the Center for Stock Return Prices value weighted portfolio over the period from July 22, 1974 to December 29, 2006 ($T = 8,192$).

persistent dynamics.

Our negative estimate of ϕ stands in contrast to the findings of Bollerslev and Mikkelsen (1996), who also find the conditional heteroskedasticity of daily stock market returns following a non-stationary, long-memory, process but with an autoregression parameter that is positive. Unlike the approach we take here of first fitting a long-memory model of the conditional variance and then adding the autoregressive term, Bollerslev and Mikkelsen first estimate a short-memory autoregressive model finding, as expected, ϕ to be very close to one. When they include the fractional differencing operator in their model it captures the non-stationary persistent behavior of the conditional heteroskedasticity in an estimate of d greater than one-half. This comes at the expense of ϕ whose value declines but continues to be positive and large relative to the unit circle. Hence, in Bollerslev and Mikkelsen (1996) the highly persistent dynamics of conditional heteroskedasticity is captured by both the fractional differencing operator and the autoregressive term.

One possible explanation for our ϕ being opposite in sign to Bollerslev and Mikkelsen (1996) is found in their finite series representation of the fractional differencing operators infinite series. By truncating the fractional differencing operator they implicitly assume shocks associated with the dropped terms have no effect on current levels of volatility. The implications of this is there will be less persistence associated with the truncated series relative to the actual fractional differencing operator. Since the persistence associated with the autoregressive term is unaffected by the truncation, the autoregressive term picks up the slack and tries to capture this lost persistence through a larger value of ϕ .

Our approach to estimating d does not approximate the fractional differencing operator. Instead, it uses the entire wavelet representation of the fractional differencing operator so that our estimator of d captures all of the persistence associated with the fractional differencing operator. Our approach leaves no leftover long-memory behavior in the volatility for the autoregressive term to pick up. The ϕ estimated with our wavelet domain approach is able to model the short term dynamics of volatility and is not contaminated by having to mop up any left over long-memory behavior lost from approximating the fractional differencing operator. With our results, we can safely say market volatility is a highly persistent, non-stationary, mean-reverting process, whose short-run dynamics consist of a back and forth, flipping dynamic.

7 Conclusion

In this paper we have designed a quick and efficient Markov chain Monte Carlo sampler of the fractional differencing and autoregressive parameters from the posterior distribution of a weakly and strictly stationary SV-FIAR model. The key contributions of our MCMC sampler is the joint estimation of the autoregressive and fractional differencing parameter from within the wavelet domain, the ability of the Bayesian, wavelet domain, estimator to handle fractionally integrated volatility that are either weakly stationary processes with finite variance, or strictly stationary processes whose variance is infinite but still mean-reverting. Simulations show that the proposed MCMC sampler is a reliable estimator of the SV-FIAR model's unknown parameters. Applying the SV-FIAR model's Bayesian, wavelet domain estimator to CRSP market returns shows the long-memory observed in the volatility of stock market returns is of the highly persistent, infinite variance, but mean-reverting type that comes from the fractional differencing parameter being larger than $1/2$.

References

- Andersen, T., Bollerslev, T., Diebold, F.X., and H. Ebens (2001) "The distribution of realized stock return volatility," *Journal of Financial Economics* 61, 43-76.
- Andersen, T., Bollerslev, T., Diebold, F.X., and P. Labys (2001) "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association* 96, 42-55.
- Andersen, T., Bollerslev, T., Diebold, F.X. and P. Labys (2003) "Modeling and forecasting realized volatility," *Econometrica* 71, 529-626.
- Arteche, J. (2004) "Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models," *Journal of Econometrics* 119, 131-154.
- Bae, S.K, Jensen, M.J, and Murdock, S.G. (2005) "Long-run neutrality in a fractionally integrated model," *Journal of Macroeconomics* 27, 257-274.
- Baillie, R.T. (1996) "Long-memory processes and fractional integration in econometrics," *Journal of Econometrics*, 73, 5-59.
- Baillie, R.T., Bollerslev, T., and H.O. Mikkelsen (1996) "Fractionally integrated generalized autoregressive conditional hetroskedasticity," *Journal of Econometrics*, 74, 3-30.
- Beran, J. (1994) *Statistics for Long-Memory Processes*, Chapman & Hall, New York.
- Bollerslev, T. and H.O. Mikkelsen (1996) "Modeling and pricing long memory in stock market volatility," *Journal of Econometrics*, 73, 151-184.
- Bollerslev, T. and H.O. Mikkelsen (1999) "Long-term equity anticipation securities and

- stock market dynamics,” *Journal of Econometrics*, 92, 75-99.
- Breidt, F.J., Crato, N. and P. de Lima (1998) “On the detection and estimation of long-memory in stochastic volatility,” *Journal of Econometrics*, 83, 325-348.
- Brockwell, P. J., and R.A., Davis, (1993) *Time Series: Theory and Methods*, 2nd ed., Springer, New York.
- Chan, N.H. and W. Palma (1998) “State space modeling of long-memory processes,” *Annals of Statistics*, 26, 719-740.
- Cheung, Y.W, and Lai, K.S. (1993) “A fractional cointegration analysis of purchasing power parity,” *Journal of Business and Economic Statistics* 11, 103-112.
- Chib, S. (2001) ”Markov chain Monte Carlo methods: Computation and inference,” in J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Elsevier Science, Amsterdam, 3569-3649.
- Chib, S. and E. Greenberg (1995) “Understanding the Metropolis-Hastings Algorithm,” *American Statistician*, 49, 327-335.
- Chib, S. and E. Greenberg (1996) “Markov chain Monte Carlo simulation methods in econometrics,” *Econometric Theory*, 12, 409-431.
- Chib, S. and E. Greenberg (1998) “Analysis of multivariate probit models,” *Biometrika*, 85, 347-361.
- Chib, S., Nardari, F. and N. Shephard (2002) “Markov chain Monte Carlo methods for stochastic volatility models,” *Journal of Econometrics*, 108, 281-316.
- Daubechies, I. (1988) “Orthonormal bases of compactly supported wavelets,” *Communication in Pure Applied Mathematics*, 41, 909-996.
- Daubechies, I. (1992) *Ten Lectures on Wavelets*, SIAM, Philadelphia.
- Deo, R.S. and C.M. Hurvich (2001) “On the log periodogram regression estimator of the memory parameter in long-memory stochastic volatility models,” *Econometric Theory*, 17, 686-710.
- Ding, Z., Granger, C.W.J., and R.F. Engle (1993) “A long memory property of stock market returns and a new model,” *Journal of Empirical Finance*, 1, 83-106.
- Frederiksen, P. and M.Ø. Nielsen (2008) “Bias-reduced estimation of long-memory stochastic volatility,” *Journal of Financial Econometrics* 6, 496-512.
- Fuller, W.A. (1996) *Introduction to Statistical Time Series* (2nd ed.), Wiley, New York.
- Gallegati, M. and W. Simmler (2014) *Wavelet Applications in Economics and Finance*, Springer, New York.
- Gelfand, A.E., and A.F.M. Smith (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 J.M. Bernardo et al., eds., *Bayesian Statistics*, 4 Oxford University, New York, NY, 169-193.
- Gradshteyn, I.S., and I.M. Ryzhik. (1994) *Table of Integrals, Series, and Products*, translated from the Russian. Translation fifth edited by A. Jeffrey, Academic Press, San Diego.
- Granger, C.W.J., and Z. Ding (1996) “Varieties of long memory models,” *Journal of Econometrics*, 73, 61-77.
- Granger, C.W.J. and R. Joyeux (1980) “An introduction to long-memory time series models and fractional differencing,” *Journal of Time Series Analysis*, 4, 221-237.
- Harvey, A.C. (2011) “Long-memory in stochastic volatility” in J. Knight, and S. Sachell, eds., *Forecasting Volatility in the Financial Market*, Butterworth-Heineman, London.
- Harvey, A.C., and N. Shephard (1996) “Estimation of an asymmetric stochastic volatility model for asset returns,” *Journal of Business and Economic Statistics*, 14, 429-434.
- Hirano, K. (2002) “Semiparametric Bayesian inference in autoregressive panel data models,” *Econometrica*, 70, 781-799.
- Hosking, J.R.M. (1981) “Fractional differencing,” *Biometrika*, 68, 165-176.
- Hosking, J.R.M. (1984) “Modeling persistence in hydrological time series using fractional differencing,” *Water Resources Research*, 20, 1898-1908.
- Hurvich, C.M., Moulines, E., and Soulier, P. (2005) “Estimating long memory in volatility,” *Econometrica* 73, 1283-1328.
- Hurvich, C.M., and Ray, B.K. (2003) “The local Whittle estimator of long-memory stochastic volatility,” *Journal of Financial Econometrics* 1, 445-470.
- In, F., and Kim, S. (2013) *An Introduction to Wavelet Theory in Finance: A Wavelet Multiscale Approach*, World Scientific Publishing, Singapore.
- Jensen, M.J. (1999) “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, 361-387.
- Jensen, M.J. (2004) “Semiparametric Bayesian inference of long-memory stochastic volatility,” *Journal of Time Series Analysis*, 25, 895-922.
- Jensen, M.J. and J.M. Maheu (2010) “Bayesian semiparametric stochastic volatility modeling,” *Journal of Econometrics*, 157, 306-316.

- Kim, S., Shephard, N. and S. Chib (1998) “Stochastic volatility: Likelihood inference and comparison with ARCH models,” *Review of Economic Studies*, 65, 361-393.
- Lai, M.J. (1995) “On the digital filter associated with Daubechies’ wavelets,” *IEEE Transactions on Signal Processing*, 43, 2203-2205.
- Mallat, S. (1999) *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, New York.
- McCoy, E.J. and A.T. Walden (1996) “Wavelet analysis and synthesis of stationary long-memory processes,” *Journal of Computational and Graphical Statistics*, 5, 26-56.
- Nelson, D.B. (1991) “Conditional hetroskedasticity in asset pricing: A new approach,” *Econometrica*, 59, 347-370.
- Ogden, R.T. (1997) “On preconditioning the data for the wavelet transform when the sample size is not a power of two,” *Communication in Statistics B* 26, 267-285.
- Percival, D.B. (1995) “On estimation of the wavelet variance,” *Biometrika*, 82, 619-631.
- Percival, D.B. and A.T. Walden (1998) *Spectral Analysis for Physical Applications*, Cambridge University Press, Cambridge, UK.
- Percival, D.B. and A.T. Walden (2000) *Wavelet Methods for Time Series Analysis*, Cambridge University Press, Cambridge, U.K.
- Shephard, N. (2005) *Stochastic Volatility*, Oxford University Press, Oxford, U.K.
- Strang, G., and T. Nguyen (1996) *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Boston.
- Tanner, M.A. and W.H. Wong (1987) “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association*, 82, 528-550.
- Tierney, L. (1994) “Markov chains for exploring posterior distributions,” *Annals of Statistics*, 21, 1701-1762.
- Whitcher, B., Guttorp, P., and D.B. Percival (2000), “Wavelet analysis of covariance with application to atmospheric time series,” *Journal of Geophysical Research*, 105, 14941-14962.