Flexible modelling of dependence in volatility processes

M. Kalli and J. E. Griffin *

July 12, 2013

Abstract

This paper proposes a novel stochastic volatility model that draws from the existing literature on autoregressive stochastic volatility models, aggregation of autoregressive processes, and Bayesian nonparametric modelling to create a stochastic volatility model that can capture long range dependence. The volatility process is assumed to be the aggregate of autoregressive processes where the distribution of the autoregressive coefficients is modelled using a flexible Bayesian approach. The model provides insight into the dynamic properties of the volatility. An efficient algorithm is defined which uses recently proposed adaptive Monte Carlo methods. The proposed model is applied to the daily returns of stocks.

*Maria Kalli, Business School, Canterbury Christ Church University, North Holmes Road, Canterbury, CT1 1QU, UK. Email: maria.kalli@canterbury.ac.uk and Jim Griffin, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury CT2 7NF, UK. Email: J.E.Griffin-28@kent.ac.uk.

Keywords: Aggregation; Long-Range Dependence; MCMC; Bayesian nonparametric; Dirichlet process; Stochastic volatility

1 Introduction

Stochastic volatility (SV) models have become a popular method for modelling financial data. The SV model described in Taylor (1986) and Harvey (1998) assumes that returns, y_t, are modelled by

$$y_t = \beta \exp\{h_t/2\}\epsilon_t, \qquad t = 1, 2, \ldots, T \tag{1}$$

where ϵ_t are i.i.d. draws from some distribution (usually, taken to be normal) and $\exp\{h_t/2\}$ is the volatility on the t-th day which is assumed to follow a stochastic process. This volatility process can be thought to represent the flow of information to the market and log volatility is often assumed to follow an AR(1) process

$$h_t = \phi h_{t-1} + \eta_t, \qquad t = 1, 2, \ldots, T \tag{2}$$

where η_t is normally distributed with mean 0 and variance $\sigma^2(1-\phi^2)$. This choice of distribution for η_t results in the stationary distribution of h_t being normal with mean 0 and variance σ^2. The autoregressive coefficient of h_{t-1}, ϕ, is the persistence parameter measuring the first lag autocorrelation of h_t.

The AR(1) assumption has become standard but there are no overriding economic reasons for its choice. Empirical analyses of financial return series have suggested that the volatility process is time-varying and mean-reverting with volatility clustering, for a comprehensive description of these stylized facts (see *e.g.* Cont, 2001; Tsay, 2005;

Taylor, 2005). These analyses have also observed that the rate of decay of the sample autocorrelation function of the squared and absolute returns is much slower than would be expected with an AR(1) volatility process. The slow decay of the sample autocorrelation function has been linked to the concept of long range dependence. We are interested in the case where h_t is the aggregate of weakly stationary processes, with a clearly defined covariance function (and thus spectral density). Then long range dependence occurs when this covariance function is unsummable (see Granger, 1980). Alternatively, when h_t is a stationary process, long range dependence can be defined in terms of fractional differencing. A process has long range dependence if the spectral density $S(\omega)$ converges to $\kappa|\omega|^{-2d}$ at very low frequencies, i.e as $\omega \to 0$, where the differencing parameter $d \in (0, 1/2)$ and $\omega \in [-\pi, \pi]$.

Early evidence of the slow rate of decay of the sample autocorrelation function in financial time series can be traced back to Ding et al. (1993), De Lima and Grato (1994), and Bollerslev and Mikkelsen (1996). Ding et al. (1993) constructed a series of fractional moments using the daily returns of the S&P 500 and found very slowly decaying autocorrelations for these series, De Lima and Grato (1994) applied some long memory tests to the squared residuals of filtered daily US stock returns and rejected the null hypothesis of short memory for these returns. Bollerslev and Mikkelsen (1996) found slowly decaying autocorrelations for the absolute returns of the S&P 500 and proposed the Fractionally Intergrated GARCH (FIGARCH) as well as the Fractionally Intergrated EGARCH (FIEGARCH). In terms of modelling long range dependence in SV models (LMSV), Harvey (1998) proposed an SV model driven by fractional noise where the log volatility h_t is expressed as a simple Autoregressive Fractionally Integrated Moving Average, ARFIMA$(0, d, 0)$. Breidt et al. (1998) extended this model by expressing the log volatility, h_t as an ARFIMA(p,d,q). Generalizations of the LMSV, appear in Arteche (2004) and Hurvich et al. (2005), where the Gaussianity assumption for h_t is replaced by a linearity assumption in both cases. The authors use the results in Surgailis and Viano (2002), and confirm that under linearity for h_t and other weak assumptions, powers of absolute returns have long range dependence.

Cross-sectional aggregation of AR(1) processes was introduced in Robinson (1978) and further explored in Granger (1980) and Zaffaroni (2004). The aggregation of such processes leads to a class of models with long range dependence which differ from fractionally integrated models, where the series requires fractional differencing to achieve a stationary ARMA series.

Suppose that we have m time series $h_{i,1}, h_{i,2}, \ldots, h_{i,T}$ for $i = 1, \ldots, m$ of the form

$$h_{i,t} = \phi_i h_{i,t-1} + \eta_{i,t} \tag{3}$$

where $\eta_{i,t} \sim \mathrm{N}(0, \sigma^2(1-\phi_i^2))$ are idiosyncratic shocks, and the persistence parameter $\phi_i \overset{iid}{\sim} F_\phi$, with support on $(0,1)$. The aggregate process is

$$h_t = \frac{1}{m}\sum_{i=1}^{m} h_{i,t}, \qquad t = 1, 2, \ldots, T. \tag{4}$$

Granger (1980) studied the effect of the distributional choice of F_ϕ on the properties and the dependence structure of the aggregate. He showed that the aggregate has the spectrum of an ARMA$(m, m-1)$ if F_ϕ is discrete on region $(-1,1)$. If F_ϕ is continuous and ϕ can take on any value on some region, then the spectrum of the aggregate does not have the form of an ARMA spectrum. He observed that the behaviour of f_ϕ, the density of F_ϕ, is important when $\phi \approx 1$, because it has an effect on long range dependence. He assumed that F_ϕ is a beta distribution on $(0,1)$ with shape parameters a and b (i.e. Be(a,b)), and confirmed that the key parameter, in terms of long range dependence, is b. He showed that when $b \to \infty$ the autocovariance function of the aggregate approximates that of an ARMA process. Zaffaroni (2004) generalises the results of Granger (1980) to distributions with density $f_\phi(\phi) \propto g(\phi)(1-\phi)^b$ on $(0,1)$ and considers the limit of the process $h_t \Big/ \sqrt{\text{Var}\left[h_t\right]}$ rather than the limit of its autocorrelation function. He showed that the process is stationary if $b > 0$ but non-stationary if $b < 0$.

This paper describes a Bayesian nonparametric approach to estimating the distribution F_ϕ. Bayesian nonparametric models place a prior on an infinite dimensional parameter space and adapt their complexity to the data. A more appropriate term is infinite capacity models, emphasising the crucial property that they allow their complexity (i.e., the number of parameters) to grow as more data is observed; in contrast, finite-capacity models assume a fixed complexity. Hjort et al. (2010) is a recent book length review of Bayesian nonparametric methods. The distribution F_ϕ is assumed to be discrete which allows us to decompose the aggregate process into processes with different levels of dependence. This models the effect of uneven information flows on volatility and can be linked to the differences in effects of different types of information. Some information may have a longer lasting effect on volatility than other pieces of information. A similar approach is discussed in Griffin (2011) using continuous time non-Gaussian Ornstein-Uhlenbeck processes for the volatility. Inference is made using Markov chain Monte Carlo (MCMC) methods with a finite approximation to the well-known Dirichlet process which exploits the relationship between the Dirichlet and gamma distributions (Ishwaran and Zarepour, 2000, 2002). The offset mixture representation of the stochastic volatility model (Kim et al., 1998) allows us to jointly update the volatilities using the Forward Filtering Backwards sampling algorithm of Frühwirth-Schnatter (1994) and Carter and Kohn (1994). This combined with recently developed adaptive MCMC methodology enables us to construct an efficient MCMC algorithm for these models.

The structure of this paper is as follows: in Section 2 we describe in detail the use of Bayesian nonparametric priors in aggregation models, Section 3 describes our sampling methodology, in Section 4 we provide illustrations with both simulated and real data, Section 5 is the discussion.

2 Bayesian nonparametric inference in aggregation models

The core of Bayesian nonparametrics is placing a prior on an infinite dimensional parameter space. In our context, the parameter is a probability distribution and the prior is a stochastic process. We will define a Bayesian nonparametric model for cross-sectional aggregation models in two stages. Firstly, we construct a suitable limiting process for a cross-sectional aggregation model as the number of elements tends to infinity. Secondly, we discuss the use of a Dirichlet process prior (Ferguson, 1973) for F_ϕ, the distribution of the persistence parameter ϕ.

We use the notation $h_t(\phi, \sigma^2)$ to represent an AR(1) process with persistence parameter ϕ and stationary variance σ^2 so

$$h_t(\phi, \sigma^2) = \phi h_{t-1}(\phi, \sigma^2) + \eta_t$$

where $\eta_t \sim \mathrm{N}\left(0, \sigma^2(1-\phi^2)\right)$, and so the marginal distribution of $h_t(\phi, \sigma^2)$ is $\mathrm{N}(0, \sigma^2)$. We define the aggregate in (4) as:

Definition 1 *A finite cross-sectional aggregation (FCA) process $h_t^{(m)}$ with parameters m, σ^2 and F_ϕ is defined by*

$$h_t^{(m)} = \frac{1}{m}\sum_{i=1}^{m} h_{i,t}\left(\phi_i, \sigma^2\right), \qquad t = 1, 2, \ldots, T \tag{5}$$

where $\phi_1, \ldots, \phi_m \overset{i.i.d.}{\sim} F_\phi$.

The stationary distribution of the FCA process is $\mathrm{N}(0, \sigma^2/m)$ and the autocorrelation function, $\rho_s = \mathrm{Corr}\left(h_t^{(m)}, h_{t+s}^{(m)}\right)$, has the form $\rho_s = \mathrm{E}_{F_\phi}\left[\phi^s\right]$ if it exists.

In the introduction we stressed the importance of the persistence parameter distribution, F_ϕ, in determining the dependence structure of the FCA process. To find a suitable limit for $h_t^{(m)}$ as $m \to \infty$, we will assume that F_ϕ is discrete with an infinite number of atoms, such that

$$F_\phi = \sum_{j=1}^{\infty} w_j \delta_{\lambda_j}, \tag{6}$$

where $w_1, w_2, w_3, \ldots$ are all positive, $\sum_{j=1}^{\infty} w_j = 1$, and δ_{λ_j} is the Dirac measure placing mass 1 at location λ_j. This assumption means that no long memory is present, although arbitrary levels of long range dependence exist.

The values $\phi_1, \ldots, \phi_m$ in the FCA process are sampled from F_ϕ and, since it is a discrete distribution, each ϕ_i must take a value in $\lambda_1, \lambda_2, \ldots$ and there can be ties in these values. We let $n_j^{(m)}$ be the number of ϕ_i's which are equal to λ_j. Clearly $n_1^{(m)}, n_2^{(m)}, \ldots$ will follow an infinite dimensional multinomial distribution which depends on m. It follows that, with this choice of F_ϕ, we can write the FCA process as

$$h_t^{(m)} = \frac{1}{m}\sum_{j=1}^{\infty} h_t\left(\lambda_j, n_j^{(m)}\sigma^2\right). \tag{7}$$

The stationary distribution of $h_t^{(m)}$ is $\mathrm{N}(0, \sigma^2/m)$ and standardising $h_t^{(m)}$ gives

$$\frac{h_t^{(m)}}{\sqrt{\mathrm{Var}[h_t^{(m)}]}} = \sum_{j=1}^{\infty} h_t\left(\lambda_j, \frac{n_j^{(m)}}{m}\right).$$

Recall that the sample counts $n_1^{(m)}, n_2^{(m)}, \ldots$ follow a multinomial distribution with parameters $w_1, w_2, \ldots$ and so

$$\lim_{m\to\infty} \frac{n_i^{(m)}}{m} = w_j \text{ as } m \to \infty.$$

It follows that

$$\lim_{m\to\infty} \frac{h_t^{(m)}}{\sqrt{\text{Var}[h_t^{(m)}]}} = \sum_{j=1}^{\infty} h_t\left(\lambda_j, w_j\right) \text{ in distribution.}$$

and its stationary distribution is $\text{N}(0,1)$ since this distribution does not depend on m. It is useful to have a scaled version of this limit for our modelling purposes which we call an ICA process.

Definition 2 *An infinite cross-sectional aggregation (ICA) process $h_t^{(\infty)}$ with parameters σ^2 and F_ϕ is defined by*

$$h_t^{(\infty)} = \sum_{j=1}^{\infty} h_{j,t}\left(\lambda_j, \sigma^2\, w_j\right), \qquad t = 1, 2, \ldots, T \tag{8}$$

where

$$F_\phi = \sum_{j=1}^{\infty} w_j \delta_{\lambda_j}.$$

The stationary distribution of $h_t^{(\infty)}$ is $\text{N}(0, \sigma^2)$ and the autocorrelation function, $\rho_s = \text{Corr}\left(h_t^{(\infty)}, h_{t+s}^{(\infty)}\right)$, has the form

$$\rho_s = \sum_{j=1}^{\infty} w_j \lambda_j^s = \int \lambda^s dF_\phi(\lambda).$$

The distribution of F_ϕ which defines the ICA process has a natural interpretation as w_j is the proportion of the variation in the stationary distribution explained by the j-th process which is associated with autoregressive parameter λ_j. The spectral density, $S(\omega)$, can also be expressed as an integral with respect to F_ϕ,

$$S(\omega) = \sigma^2 \int \frac{1}{(1-\lambda^2)(1+\lambda^2 - 2\lambda\cos(\omega))} dF_\phi(\lambda)$$

where $\frac{1}{(1-\lambda^2)(1+\lambda^2-2\lambda\cos(\omega))}$ is the spectral density of an AR(1) process with persistence parameter λ and marginal variance 1.

Our focus is on the estimation of F_ϕ. To avoid making parametric assumptions about this distribution, we take a Bayesian nonparametric approach and use a Dirichlet process (DP) prior for F_ϕ. The DP prior is often used to define an infinite mixture model for density estimation by giving the mixing distribution a DP prior. Our approach is substantially different and uses the Dirichlet process as prior for the distribution F_ϕ in the ICA process model. This leads to a nonparametric approach to understanding the dynamic behaviour of the aggregate $h_t^{(\infty)}$.

The properties and theory for the DP were developed in Ferguson (1973). We say that a random distribution F with sample space Ω is distributed according to a DP if the masses placed on all partitions of Ω are Dirichlet distributed. To elaborate, let F_0 be a distribution over sample space Ω and c a positive real number, then for any measurable partition $A_1, \ldots, A_r$ of Ω the vector $(F(A_1), \ldots, F(A_r))$ is random since F is random. The Dirichlet process can be defined as

$$F \sim \text{DP}(c, F_0) \text{ if } (F(A_1), \ldots, F(A_r)) \sim \text{DP}(cF_0(A_1), \ldots, cF_0(A_r)) \tag{9}$$

for every finite measurable partition $A_1, \ldots, A_r$ of Ω.

The distribution F_0 is referred to as the centring distribution as it is the mean of the DP, that is $\text{E}[F(A_i)] = F_0(A_i)$. The positive number c is referred to as the concentration (or precision) parameter as it features in and controls the variance of the DP, that is $\text{Var}[F(A_i)] = F_0(A_i)(1-F_0(A_i))/(c+1)$. One can clearly see that the larger the value of c, the smaller the variance, and hence the DP will have more of its mass concentrated around its mean.

We now describe the priors given to the parameters of the ICA process. The distribution F_ϕ is given a Dirichlet process prior with precision parameter c and a Be$(1, b)$ centring distribution. The parameter $\sigma^2 \sim$ Ga$(c, c/\zeta)$ where Ga(α_1, α_2) represents a Gamma distribution with density

$$f(x) = \frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1 - 1} \exp\{-\alpha_2 x\}, \qquad x > 0.$$

This implies that the prior mean of σ^2 is ζ. The choice of a Dirichlet process prior for F_ϕ implies that this distribution is *a priori* discrete with probability 1 and so it is suitable for use with the ICA process. An alternative parameterization of the F_ϕ is

$$F_\phi = \sum_{j=1}^{\infty} \frac{\tau_j}{\sum_{k=1}^{\infty} \tau_k} \delta_{\lambda_j}. \tag{10}$$

where $\tau_j = \sigma^2 w_j$. The choice of prior for F_ϕ and σ^2 implies that $\tau_1, \tau_2, \ldots$ are the jumps of a Gamma process with Lévy density $\eta(x) = cx^{-1} \exp\{-c/\zeta x\}$.

A stochastic volatility with ICA volatility process (SV-ICA) model can now be defined for returns $y_1, \ldots, y_T$

$$y_t = \beta \exp\{h_t^{(\infty)}/2\} \epsilon_t$$

where $\epsilon_t \sim$ N$(0, 1)$ and $h_t^{(\infty)}$ follows an ICA process. This model allows a flexible form for the autocorrelation and helps decomposition of the volatility process dynamics. The model is completed by specifying priors for the parameters β, b, c and ζ. The parameter $\mu = \log \beta^2$ is given a vague improper prior, $p(\mu) \propto 1$. This makes the prior invariant to rescaling of the data. The hyperprior of b, the scale parameter for the distribution of λ_i, is taken to be an exponential distribution with mean $1/(\log 2)$. This implies that the prior median of b is 1 and hence places half its mass on processes with long memory and half its mass on processes with short memory. This choice avoids making strong assumptions about the dynamic properties of the time series. The parameter c controls how close F_ϕ is to the beta centring distribution. We follow Griffin and Steel (2004) by defining the prior for c through $\frac{c}{c+n_0} \sim \text{Be}(a_c, a_c)$, where n_0 is a prior sample size for F_ϕ and a_c is a precision parameter (smaller values of a_c imply less variability). The density for c is

$$p(c) = \frac{n_0^n \Gamma(2a_c)}{\Gamma(2a_c)^2} \frac{c^{a_c - 1}}{(c + n_0)^{2a_c}}.$$

We choose the values $a_c = 5$ and $n_0 = 3$ which place most of the mass on relatively small values of c implying that F_ϕ is away from the beta distribution. The parameter ζ represents the prior mean of the overall variability σ^2 of the volatility process. We represent our prior ignorance about this scale by choosing the vague prior $\zeta^{-1} \sim \text{Ga}(0.001, 0.001)$.

3 Computation

Markov chain Monte Carlo (MCMC) inference in the SV-ICA model is complicated by the presence of an infinite sum of AR(1) processes and the non-linear state space form of the model. The first problem is addressed through a finite truncation of F_ϕ and the second problem through the offset mixture representation of the SV model discussed in Kim et al. (1998), Omori et al. (2007), and Nakajima and Omori (2009).

The approximation of the Dirichlet process by finite truncations has been discussed by many authors including Neal (2000), Richardson and Green (2001), Ishwaran and Zarepour (2000), and Ishwaran and Zarepour (2002). In the latter paper the authors discuss using Dirichlet random weights to construct a finite-dimensional random probability measure with n atoms that limits to a Dirichlet process as $n \to \infty$. Their approach approximates F_ϕ by

$$F_\phi^{(n)} = \sum_{j=1}^{n} w_j^{(n)} \delta_{\lambda_j^{(n)}} \tag{11}$$

where $\left(w_1^{(n)}, \ldots, w_n^{(n)}\right) \sim \text{Dir}\left(c/n, \ldots, c/n\right)$ [1] and $\lambda_j^{(n)} \overset{iid}{\sim} \text{Be}(1, b)$. The relationship between the Dirichlet distribution and the gamma distribution allows us to further write

$$F_\phi^{(n)} = \sum_{j=1}^{n} \frac{\sigma_j^2}{\sum_{k=1}^{n} \sigma_k^2} \delta_{\lambda_j} \tag{12}$$

with $\sigma_j^2 \overset{iid}{\sim} \text{Ga}(c/n, c/\zeta)$. As well as approximating the Dirichlet process prior for F_ϕ this also implies that $\sigma^2 = \sum_{j=1}^{n} \sigma_j^2 \sim \text{Ga}(c, c/\zeta)$, replicating the prior for the infinite-dimensional model. The idea of using Gamma priors for variances to avoid model over-complexity has been explored recently in the Bayesian literature on regression models (see Caron and Doucet (2008) and Griffin and Brown (2010)). The construction of $F_\phi^{(n)}$ in (11) is intuitive.

Combining the finite truncation $F_\phi^{(n)}$ of F_ϕ with the SV-ICA model leads to the model (where we drop the dependence of the model on the value of n for notational simplicity)

$$y_t = \beta \exp\{h_t/2\}\epsilon_t, \qquad h_t = \sum_{j=1}^{n} h_{j,t}^\star$$

and

$$h_{j,t}^\star = \lambda_j h_{j,t-1}^\star + \eta_{j,t}^\star, \qquad j = 1, \ldots, n$$

where $\eta_{j,t}^\star \sim \text{N}(0, \sigma_j^2(1-\lambda_j^2))$ and $\sigma_j^2 \sim \text{Ga}(c/n, c/\zeta)$.

We proceed to inference using the linearized form of the stochastic volatility model for MCMC (Kim et al., 1998; Omori et al., 2007; Nakajima and Omori, 2009). This approach works with $y_t^* = \log(y_t^2 + \xi)$ where the offset value ξ is introduced to avoid y_t^* from being very negative or even undefined, due to zero returns, see Fuller (1996). It follows that y_t^* can be expressed as a linear function of h_t,

$$y_t^* = \mu + h_t + \epsilon_t^*, \qquad t = 1, \ldots, T \tag{13}$$

[1] $\text{Dir}(\alpha_1, \alpha_2, \ldots, \alpha_{k+1})$ represents the k-dimensional Dirichlet distribution

where $\mu = \log(\beta^2)$, $\epsilon_t^* = \log(\epsilon_t^2)$ and ϵ_t^* follows a $\log \chi^2$ distribution. The $\log \chi^2$ distribution can be approximated by a mixture of 7 normal distributions

$$p(\epsilon_t^*) = \sum_{i=1}^{7} p_i \, \mathrm{N}(\nu_i, \psi_i^2).$$

Introducing allocation variables $s_1, s_2, \ldots, s_T$ where $p(s_t = j) = p_j$ leads to the conditional distribution $\epsilon_t^* | s_t \sim \mathrm{N}(\nu_{s_t}, \psi_{s_t}^2)$.

The SV-ICA model with the finite truncation of F_ϕ and the linearized form for the return defines a Gaussian dynamic linear model conditional on $s_1, s_2, \ldots, s_T$ and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$,

$$\begin{aligned} y_t^* &= \mu + \mathbf{1} H_t + \epsilon_t^* \\ H_t &= \Lambda H_{t-1} + \eta_t^\star \end{aligned} \tag{14}$$

where $\mathbf{1}$ is a $(1 \times n)$ dimensional vector of 1's,

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix},$$

and

$$H_t = \begin{pmatrix} h_{1,t}^\star \\ h_{2,t}^\star \\ \vdots \\ h_{n,t}^\star \end{pmatrix}$$

$\epsilon_t^* | s_t \sim \mathrm{N}(\nu_{s_t}, \psi_{s_t}^2)$, and $\eta_t^\star$ follows a zero mean normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2(1-\lambda_1^2) & & & 0 \\ & \sigma_2^2(1-\lambda_2^2) & & \\ & & \ddots & \\ 0 & & & \sigma_n^2(1-\lambda_n^2) \end{pmatrix},$$

$$\sigma_j^2 \sim \text{Ga}\left(\frac{c}{n}, \frac{c}{\zeta}\right), \qquad \lambda_j \sim \text{Be}(1, b),$$

$$p(\mu) \propto 1 \text{ and } \zeta^{-1} \sim \text{Ga}(\alpha_\zeta, \beta_\zeta).$$

The hyperparameters are $\theta = (\sigma_1^2, \dots, \sigma_n^2, s_1, \dots, s_T, \lambda_1, \dots, \lambda_n, \mu, b, c, \zeta)$ and let $\sigma^2 = \sum_{j=1}^n \sigma_j^2$. The model for y_t^* and h_t is a linear, Gaussian state-space model and so the marginal likelihood $p(y_1^*, \dots, y_T^*|\theta)$ can be calculated analytically using the Kalman filter and θ updated using Metropolis-Hastings random walk steps. This approach, however, has some problems. The value of n will often be large and the number of computations needed to calculate the marginal likelihood will be $O(n^3)$ (due to the matrix inversions involved). This makes this scheme only computationally feasible when n is small (say, less than 10). It would also be difficult to propose a high-dimensional vector of new parameter values. On the other hand, the parameters can be sampled using a Gibbs sampler by simulating the hidden states $h_{i,t}$ using a Forward Filtering Backward Sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). However, this sampler is known to mix slowly when λ_j becomes close to 1.

We propose an algorithm which combines both samplers. The parameters are updated using the Gibbs sampler. Additional steps are added which choose $k < n$ components and update parameters conditional on the states of the other components while marginalising over the states in the k chosen components (we use $k = 4$ in our examples). This set up allows us to improve the poor mixing associated with the Gibbs sampler whilst controlling the computational cost. The sampler uses adaptive proposals, which are allowed to depend on previous values of the sampler, in a Metropolis-within-Gibbs scheme. These schemes destroy the Markovian structure of the sampler and so special theory must be developed to prove convergence of these algorithms to the correct stationary distribution. Convergence of Metropolis-within-Gibbs scheme is discussed in Haario et al. (2005), Roberts and Rosenthal (2009) and Latuszynski et al. (2013). In particular, we assume that the parameters $\sigma_1^2, \ldots, \sigma_n^2$, c and b are truncated at large values, which ensures uniform ergodicity of the chain and so ergodicity of the chain follows from Theorem 5.5 of Latuszynski et al. (2013). It will be useful to define $x^{(m)}$ to be the value of the parameter x at the start of the m-th iteration. The steps of the sampler are:

Updating $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$

The full conditional distribution of σ_j^2 is

$$\text{GIG}\left(\frac{c}{n} - \frac{1}{2}T, h_{1,j}^2 + \frac{\sum_{t=2}^{T}\left(h_{t,j} - \lambda_j h_{t-1,j}\right)^2}{1-\lambda_j^2}, 2\frac{c}{\zeta}\right)$$

where GIG(a_1, a_2, a_3) represents the Generalized Inverse Gaussian distribution which has density proportional to

$$y^{a_1-1}\exp\left\{-\frac{1}{2}\left(\frac{a_2}{y} + a_3 y\right)\right\}.$$

Updating $s_1, \ldots, s_T$

The full conditional distribution of s_t is

$$p(s_t = j) \propto p_j \psi_j^{-1} \exp\left\{-\frac{1}{2}\frac{(y_t - h_t - \nu_j)^2}{\psi_j^2}\right\}, \qquad j = 1, \ldots, 7$$

Updating $\lambda_1, \ldots, \lambda_n$

The full conditional distribution of λ_j is proportional to

$$(1-\lambda_j)^{b-1}(1-\lambda_j^2)^{-(T-1)/2} \exp\left\{-\frac{1}{2}\frac{\sum_{t=2}^{T}(h_{t,j}-\lambda_j h_{t-1,j})^2}{\sigma_j^2(1-\lambda_j^2)}\right\}$$

which is updated using a random walk on the logit scale (i.e. the proposal value at the m-th iteration λ_j' is simulated as $\log\lambda_j' - \log(1-\lambda_j') \sim \mathrm{N}(\log\lambda_j^{(m)} - \log(1-\lambda_j^{(m)}), \sigma_{\lambda,j}^{2\,(m)})$). The tuning parameter $\sigma_{\lambda,j}^{2\,(m)}$ is updated in the sampler so that the acceptance rate converges to 0.3 (following the advice of Roberts and Rosenthal (2009)) using the method of Atchadé and Rosenthal (2005). This involves setting

$$\sigma_{\lambda,j}^{2\,(m+1)} = \rho\left(\sigma_{\lambda,j}^{2\,(m)} + m^{-0.6}\,(\alpha - 0.3)\right)$$

where α is the acceptance probability from the update at the m-th iteration and

$$\rho(x) = \begin{cases} L & \text{if } x \le L, \\ x & \text{if } L < x < U, \\ U & \text{if } x \ge U. \end{cases}$$

where L and U are user-defined values which are respectively very small and very large. The function ρ is introduced to ensure the stability of the adaptive method.

Updating μ

The full conditional of μ is $\mathrm{N}\left(\bar{y}^\star, \sigma^{\star\,2}\right)$ where $\bar{y}^\star = \frac{\sum_{t=1}^{T}\left(y_t^\star - \sum_{j=1}^{n} h_{t,j}^\star\right)/\psi_{s_t}^2}{\sum_{t=1}^{T} 1/\psi_{s_t}^2}$ and $\sigma^{\star\,2} = \frac{1}{\sum_{t=1}^{T} 1/\psi_{s_t}^2}$.

Updating ζ

The full conditional distribution of ζ^{-1} is $\mathrm{Ga}(c + \alpha_\zeta, c\sum_{j=1}^{n}\sigma_j^2 + \beta_\zeta)$.

Updating b and c

The full conditional distribution of b is proportional to

$$p(b)b^n\left[\prod_{j=1}^{n}(1-\lambda_j)\right]^b$$

and the full conditional distribution of c is proportional to

$$p(c)\frac{(c/\zeta)^n}{\Gamma(c/n)^n}\left(\prod_{j=1}^{n}\sigma_j^2\right)^{c/n}\exp\left\{-\frac{c}{\zeta}\sum_{j=1}^{n}\sigma_j^2\right\}.$$

These parameter can be updated one-at-a-time using a Metropolis-Hastings random walk with normal increments on the log scale. For example, b is updated using a proposal of the form $\log b' \sim \mathrm{N}\left(\log b, \sigma_b^{2\,(m)}\right)$. The variances of the proposal are updated using the adaptive scheme described in the update of $\lambda_1, \ldots, \lambda_n$.

Acceleration steps

At the m-th iteration, k components are selected without replacement with the probability of choosing the j-th component being proportional to $5/n + \sum_{k=1}^{m}\sigma_j^{2\,(k)}$. Let the indices of the chosen components be $\mathcal{I}$. We define $h_{\mathcal{I}} = \{h_{t,j}|j \in \mathcal{I}\}$ and $h_{-\mathcal{I}} = \{h_{t,j}|j \notin \mathcal{I}\}$. In each step, parameters are updated marginalizing over $h_{\mathcal{I}}$ and conditioning on $h_{-\mathcal{I}}$ (and any other parameters). This leads to the likelihood $p(y_1^*, \ldots, y_T^*|h_{-\mathcal{I}}, \theta)$ which can be efficiently computed using the Kalman filter. We also define $\lambda_{\mathcal{I}} = \{\lambda_i|i \in \mathcal{I}\}$ and $\sigma_{\mathcal{I}}^2 = \{\sigma_i^2|i \in \mathcal{I}\}$. The steps involve updating

Acceleration step 1: Updating μ

The parameter μ is updated using a Metropolis-Hastings random walk with a normal increment whose variance is tuned using the adaptive algorithm of Atchadé and Rosenthal (2005), described in the update of $\lambda_1, \ldots, \lambda_n$.

Acceleration step 2: Updating $\sigma^2_{\mathcal{I}}$

The elements of $\sigma^2_{\mathcal{I}}$ are updated in a block conditional on $\sum_{j\in\mathcal{I}} \sigma^2_j$ using a Metropolis-Hastings random walk. Let $w_h = \frac{\sigma^2_{\mathcal{I}_h}}{\sum_{j\in\mathcal{I}} \sigma^2_j}$ then $w_1, \ldots, w_k \sim$ Di $\left(\frac{c}{n}, \ldots, \frac{c}{n}\right)$. Let $z_i = \log(w_{i+1}) - \log(w_1)$ for $i = 1, \ldots, k-1$ and calculate the covariance

$$C^{(2)}_{i,j} = \frac{1}{m-2}\left[\sum_{h=1}^{m-1} \log w_i^{(h)} \log w_j^{(h)} - \frac{\sum_{h=1}^{m-1} \log w_i^{(h)} \sum_{h=1}^{m-1} \log w_j^{(h)}}{m-1}\right]$$

from the previous samples. A proposed value z' can be generated using a random walk proposal where the increment is multivariate normal with the covariance matrix being a scaled version of $C^{(2)}$. This can be used to derive proposed values w' and $\sigma^{2'}_{\mathcal{I}}$. The Metropolis-Hastings acceptance probability is

$$\max\left\{1, \frac{p(y_1^*, \ldots, y_T^*|h_{-\mathcal{I}}, \theta') \prod_{j\in\mathcal{I}} {w'_j}^{c/n-1}|M'|}{p(y_1^*, \ldots, y_T^*|h_{-\mathcal{I}}, \theta) \prod_{j\in\mathcal{I}} {w_j}^{c/n-1}|M|}\right\}$$

where M' and M are $(k-1)\times(k-1)$-dimensional matrix with entries

$$M'_{ij} = \begin{cases} \frac{1}{w'_{\mathcal{I}_1}} & i \neq j \\ \frac{1}{w'_{\mathcal{I}_1}} + \frac{1}{w'_{\mathcal{I}_{i-1}}} & i = j \end{cases} \quad \text{and } M_{ij} = \begin{cases} \frac{1}{w_{\mathcal{I}_1}} & i \neq j \\ \frac{1}{w_{\mathcal{I}_1}} + \frac{1}{w_{\mathcal{I}_{i-1}}} & i = j \end{cases}.$$

Acceleration step 3: Updating λ

Each element of $\lambda_{\mathcal{I}}$ is updated separately using a Metropolis-Hastings random walk on the logit scale. Let $j \in \mathcal{I}$ then the proposed value is $\log\lambda'_j - \log(1-\lambda'_j) \sim$ N$(\log\lambda_j - \log(1-\lambda_j), \sigma^2_{f,j})$ where $\sigma^2_{f,j}$ is the variance of the increments for λ_j. The proposed value is accepted according to the Metropolis-Hastings acceptance probability

$$\max\left\{1, \frac{p(y_1^*, \ldots, y_T^*|h_{-\mathcal{I}}, \theta')\lambda'_j(1-\lambda'_j)^b}{p(y_1^*, \ldots, y_T^*|h_{-\mathcal{I}}, \theta)\lambda_j(1-\lambda_j)^b}\right\}.$$

and $\sigma^2_{f,j}$ is tuned using the adaptive algorithm of Atchadé and Rosenthal (2005).

4 Examples

We illustrate the use of the SV-ICA model on one simulated example and two real data examples: the returns of HSBC plc and Apple Inc. Our main focus is inference about the distribution F_ϕ since this represents the decomposition of the volatility process in terms of AR(1) processes with different first-lag dependences.

The SV-ICA model fit by our MCMC sampler uses the truncation $F_\phi^{(n)}$ (see equations (11) and (12)), of F_ϕ. As $n \to \infty$ the posterior of the SV-ICA model using $F_\phi^{(n)}$ should converge to the posterior of the SV-ICA model using F_ϕ. To assess the convergence of $F_\phi^{(n)}$ to F_ϕ we inspect the posterior expectation of $F_\phi^{(n)}([0, x])$ for $x \in (0, 1)$, the cumulative distribution function, at different values of n. Informally, we can be happy that the posterior is converging if there are only small changes in the posterior summaries (posterior expectation and 95% credible interval) for n larger than some n_0. We have found that running $n = 30$, $n = 50$ and $n = 70$ were sufficient to judge convergence in our three examples.

To gain more insight into our decomposition of the persistence in returns, we calculate the proportion of processes for which the dependence is small by lag κ. We choose to define this measure by

$$\gamma_\kappa = F_\phi^{(n)}\left(\{\lambda \,|\lambda^\kappa < \varepsilon\}\right) \tag{15}$$

for some small value ε (we take $\varepsilon = 0.01$).

The parameters c and ζ have clear interpretation in the SV-ICA model. Recall from Section 2 that the concentration parameter c controls the variability of the Dirichlet process. This is also the case for $F_\phi^{(n)}$. The variance of $F_\phi^{(n)}$, Var $\left[F_\phi^{(n)}\right]$, is affected by changes in the value of c. It becomes smaller as the value of c increases, hinting that $F_\phi^{(n)}$ is close to its centering distribution. This gives a simple interpretation of c.

Since $\sigma^2 \sim$ Ga$(c, c/\zeta)$, the value of ζ is of interest as it directly affects the scale of the distribution. For both c and ζ we present tables of their posterior medians together with the 95% credible intervals.

All MCMC algorithms were run for 80 000 iterations with 20 000 discarded as burn-in which was sufficient for convergence of the Markov chain. The offset parameter was set to $\xi = 10^{-5}$.

4.1 Simulated data

The simulated data series are based on the stochastic volatility model of equation (1) and the volatility process, h_t, is the aggregation of 50 AR(1) processes, that is

$$y_t = \beta \exp\{h_t/2\}\epsilon_t, \qquad h_t = \frac{1}{50}\sum_{i=1}^{50} h_{i,t}, \qquad t = 1, \ldots, T$$

where $h_{i,t}$ is an AR(1) process with persistence parameter ϕ_i and increments $\eta_{i,t} \sim$ N$(0, \sigma^2(1-\phi_i^2))$. The parameter $\beta = 1$, $\sigma^2 = 1$ and $\phi_i \overset{iid}{\sim}$ Be$(1, 0.7)$. The length of the time series is $T = 2000$.

The plots of the posterior expectation of $F_\phi^{(n)}$ together with the 95% credible interval, for each value of n are shown in Figure 1. Following referee advice we also display the estimate of the posterior density of ϕ under a simple SV model for comparison.

The posterior expectation of $F^{(n)}$ becomes closer to the true generating cumulative distribution function as n increases; the credible interval also becomes narrower. A closer look at the plots reveals a bi-modal distribution for $n = 30$ which changes to unimodal for $n = 50$ and $n = 70$. These results indicate that $F_\phi^{(n)}$ begins to converge around $n = 50$, which is not surprising since the data are generated with 50 underlining AR(1) processes. The median of ϕ for $n = 50$ and $n = 70$ appears to be within $(0.7, 0.8)$.

The posterior density of ϕ under the simple SV model also hints to a median within $(0.7, 0.8)$.

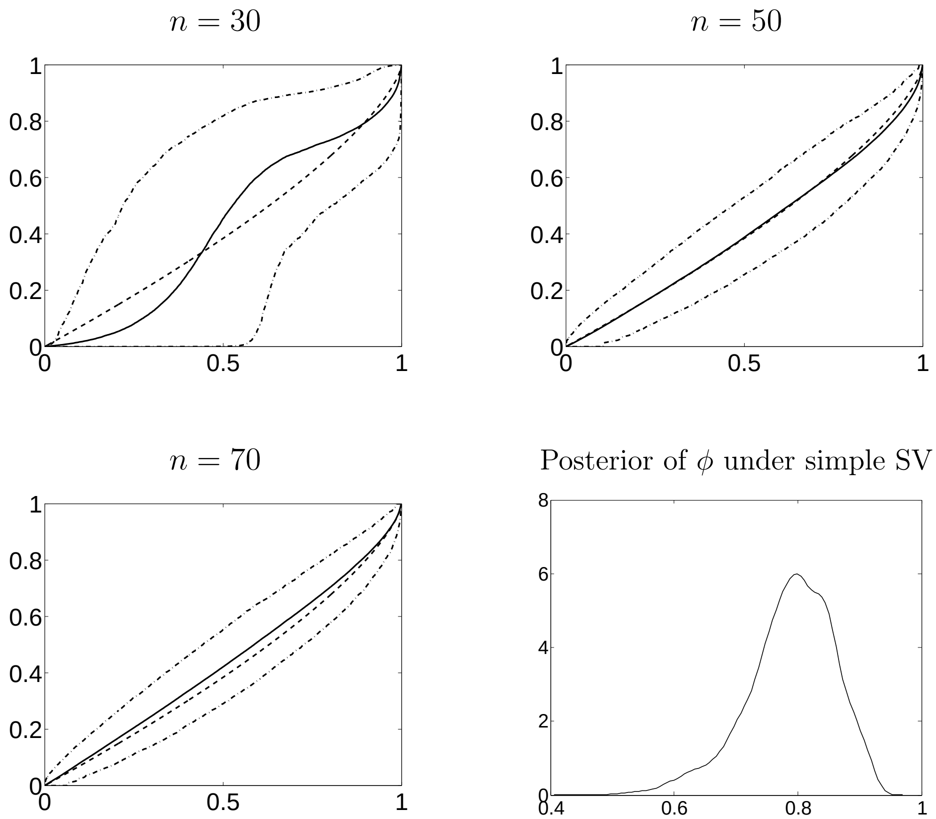

Figure 1: Simulated data: The true cumulative distribution function (dashed line) with the posterior expectation of $F_\phi^{(n)}$ (solid line), and 95% credible interval (dot-dashed lines) for $n = 30$, $n = 50$, and $n = 70$. The x-axis is ϕ and the y-axis is the posterior expectation of $F_\phi^{(n)}$. The bottom, right-hand graph is the posterior density of ϕ under the simple SV model.

Table 1 displays the posterior median and 95% credible intervals for the precision parameter c and ζ. The median of c increases substantially as n increases. From 4.4 at $n = 30$ to 934.6 at $n = 50$. Given the properties of the Dirichlet process, the Var $\left[F_\phi^{(n)}\right]$ becomes smaller and smaller showing that $F_\phi^{(n)}$ is close to the centring distribution. This indicates the correct data generating mechanism. The median of ζ

changes from 1.1 to 1.2 when n increases to 50 and settles at 1.2 even with $n = 70$, which is consistent with our choice of σ^2.

	$n = 30$	$n = 50$	$n = 70$
c	4.4 (2.9, 6.8)	934.6 (377.4, 2619.1)	2168.4 (1034.1, 5695.2)
ζ	1.1 (0.5, 3.7)	1.2 (1.0, 1.5)	1.2 (1.0, 1.4)

Table 1: Simulated data: The posterior median and 95% credible interval for c and ζ for $n = 30$, $n = 50$, and $n = 70$.

4.2 Real data

The real data series used to illustrate our method are: the daily returns of HSBC plc from May 16^{th} 2000 to July 14^{th} 2010, and the daily returns of Apple Inc. from January 1^{st} 2000 to July 26^{th} 2010.

The plots of these returns are shown in Figure 2. The returns of HSBC appear more volatile than those of Apple. There is a period of relatively low volatility for HSBC from March 2003 till February 2008, with two periods of high volatility, one from October 2000 to November 2002 (the biggest dip in returns occurring around 9/11), and another from January 2008 to March 2009. This latter period is due to the collapse of the banking sector following the collapse of the US subprime mortgage market. In March 2009 HSCB incurred a loss of $17 billion due to its exposure to US mortgage market. The returns of Apple Inc are not as volatile as those of HSBC plc. It operates in the technology sector where large spikes in returns are related to the effectiveness of the company in keeping up with the speed of technical advances and in launching innovative products. The period of high volatility around September 2000 is due to the introduction of the OS X operating system and the iMac which led to a 30% increase in revenue. The next high volatility period around September 2008 is due to the increase in net revenue by 50% caused by the introduction of the iPhone 3G earlier in the year.

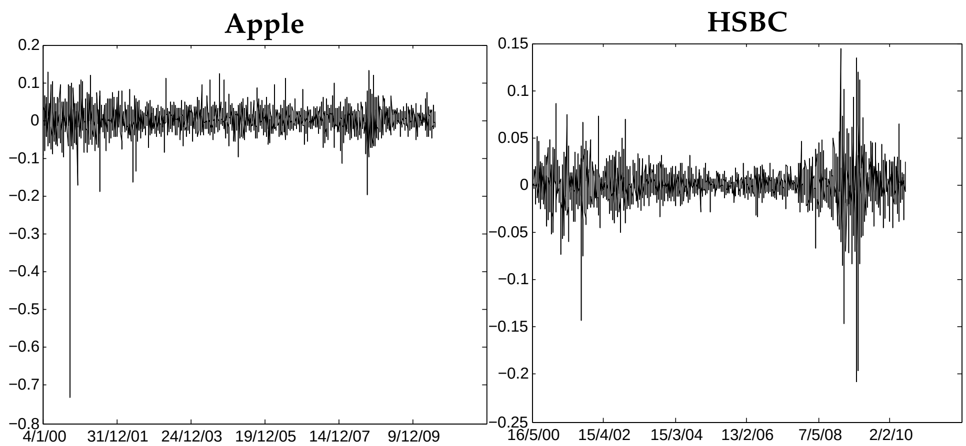

Figure 2: The returns for Apple and HSBC

As with our simulated example we focus our attention on the convergence of $F_\phi^{(n)}$ as n increases. For each data series and value of n, we provide plots of the posterior expectation of $F_\phi^{(n)}$ together with 95% credible intervals, as well as a plot of the estimated density of ϕ under the simple SV model. We also provide the value of the posterior median (together with 95% credible intervals) of c and ζ.

The results for HSBC are shown in Figure 3, and Tables 2 and 3. The plots in Figure 3 show that the posterior expectation of $F_\phi^{(n)}$ and its 95% credible intervals are similar across the three values of n considered. Focusing on $n = 70$, we observe that much of its mass is placed close to one. This is similar to the fit of a simple SV model with a single AR(1) process for h_t, the volatility process. We fitted the simple SV model using the priors of Kim et al. (1998) and found that ϕ has a posterior median of 0.984 with a 95% credible interval (0.976, 0.992); the last plot in Figure 3 confirms these findings. However, from our plots of the posterior expectation of $F_\phi^{(n)}$ we can see that there is also mass at much smaller and much larger values of ϕ. This implies that values of ϕ_i are more spread out within the interval $(0.5, 1)$, compared to what the simple SV model suggests.

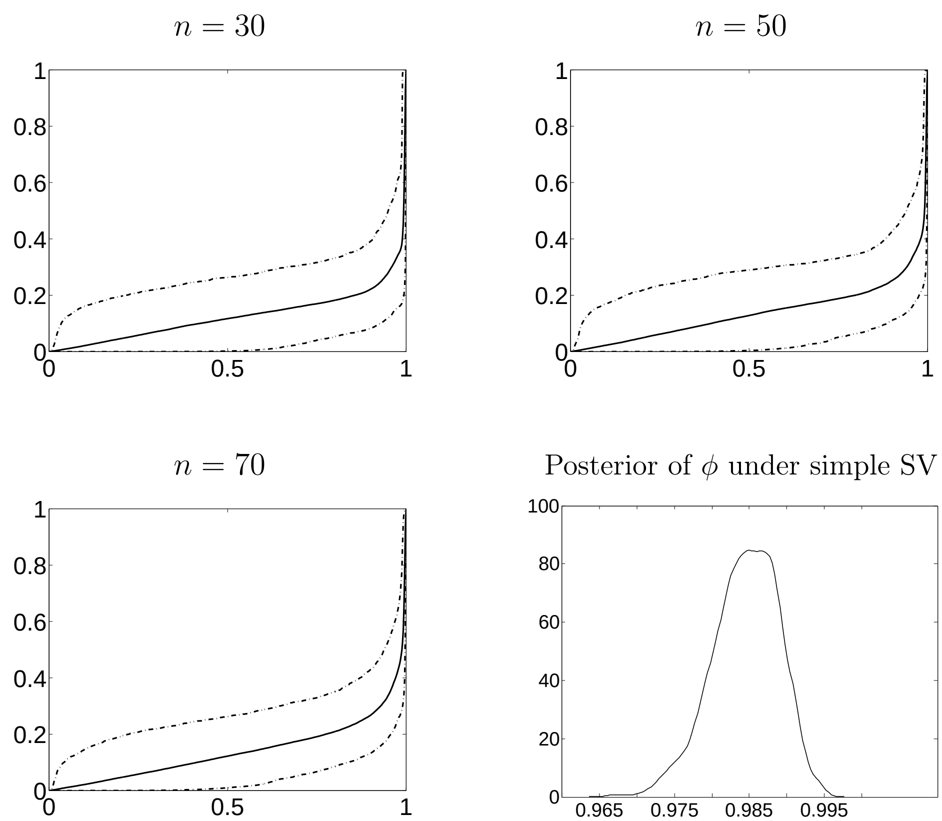

Figure 3: HSBC data: Posterior expectation of $F_\phi^{(n)}$ (solid line) with 95% credible interval (dot-dashed lines) for $n = 30$, $n = 50$, and $n = 70$. The x-axis is ϕ and the y-axis is the posterior expectation of $F_\phi^{(n)}$. The bottom, right-hand graph is the posterior density of ϕ under the simple SV model.

Table 2 provides values for the posterior median and the 95% credible intervals for the precision parameter c and ζ for the three different values of n. The median values for c, though they increase as the value of n increases, they are relatively small in comparison to the values we had in our simulated example. This finding means that the Var $\left[F_\phi^{(n)}\right]$ is large, which provides evidence to support the argument that the

	$n = 30$	$n = 50$	$n = 70$
c	5.3 (3.3, 8.4)	9.1 (6.3, 13.4)	14.3 (10.5, 19.6)
ζ	1.2 (0.5, 3.8)	1.0 (0.5, 2.5)	0.9 (0.5, 1.9)

Table 2: HSBC data: The posterior median and 95% credible interval for c and ζ for $n = 30$, $n = 50$, and $n = 70$.

beta distribution (which has been the popular choice for the distribution of ϕ) is not a particularly good fit for the distribution of ϕ for the HSBC returns, for the sample period we have analysed.

1 week	2 weeks	8 weeks	1/2 year	1 year	2 years	5 years
0.10	0.16	0.26	0.38	0.45	0.54	0.81

Table 3: HSBC data: Values of γ_κ at various lags when $n = 70$

Table 3 displays the values of γ_κ for various values of κ when $n = 70$. Recall that γ_κ is the proportion of processes with small levels of dependence less than 0.01 after k lags. The lags are displayed in terms of trading weeks and trading years [2]. The first entry for γ_κ in Table 3 indicates that 10% of the variation in volatility is explained by processes which decay after 1 week (decay quickly). Moving along the table we see that 45% of the variation is explained by processes that decay after one year. The posterior median estimate of the persistence parameter for the simple AR(1) model suggests that autocorrelation falls below 0.01 by the 286^{th} lag. This is roughly a little over one trading year. This is an interesting point because under our model we observe that a higher proportion of the variation in volatility is placed on processes that take

[2] A trading week has 5 days and a trading year approximately 252.

two to five years to decay. This is clearly seen in Table 2 where the autocorrelation of 19% of processes has not decayed below 0.01 after 5 years, providing evidence of very long persistence in the data.

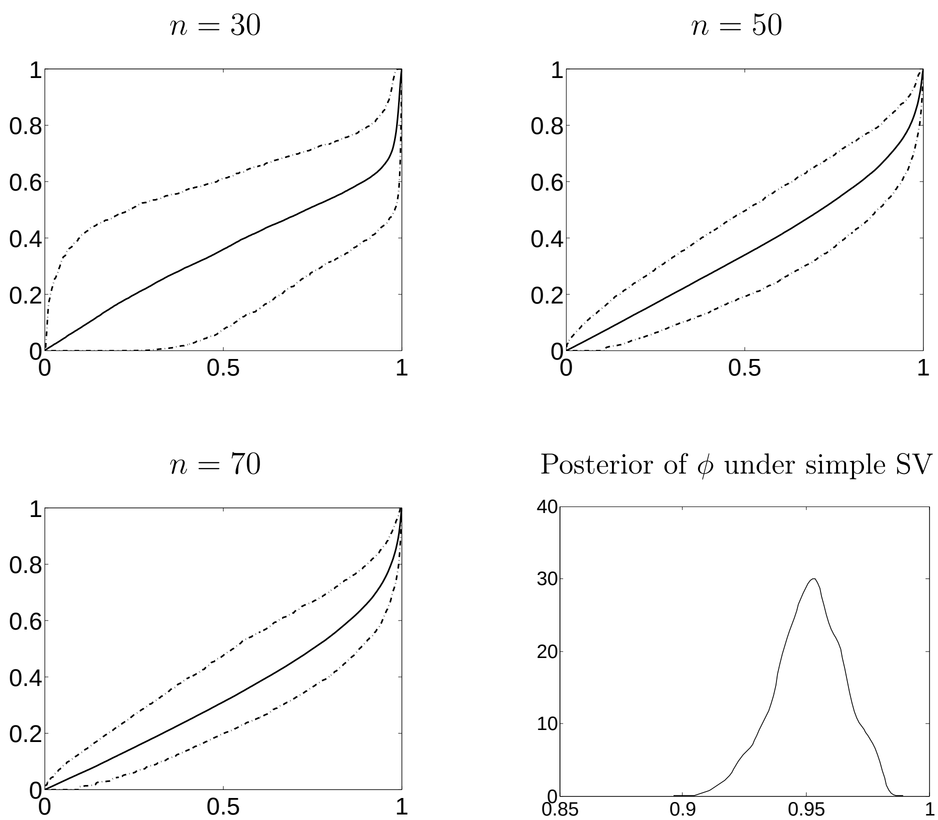

Figure 4: Apple data: Posterior expectation of $F_\phi^{(n)}$ (solid line) with 95% credible interval (dot-dashed lines) for $n = 30$, $n = 50$, and $n = 70$. The x-axis is ϕ and the y-axis is the posterior expectation of $F_\phi^{(n)}$. The bottom, right-hand graph is the posterior density of ϕ under the simple SV model.

The results for Apple are displayed in Figure 4, Table 5, and Table 4. The plots of the posterior expectation of $F_\phi^{(n)}$ in Figure 4 are very similar for all three values of n. The 95% credible interval for $n = 30$, however, is wider than those for $n = 50$ and $n = 70$ (which have similar credible intervals). This implies that $F_\phi^{(n)}$ converges around $n = 50$. Regardless of the value of n all of these plots imply that $F_\phi^{(n)}$ for Apple is quite different to that of HSBC. With HSBC more mass was placed at higher values of ϕ whereas with Apple we see that much more mass is placed at smaller values of ϕ. This difference becomes clearer when we focus on the values of γ_κ, based on $n = 70$, displayed in Table 5. The values of γ_κ are larger for all lags when compared to HSBC. In Apple's case 85% of the variation in volatility is explained by processes with autocorrelation decaying below 0.01 before one year. In contrast, this occurs after five years (or more) with the HSBC returns series. With Apple the autocorrelation of only 4% of the processes has decayed below 0.01 after five years. This implies that the behaviour of persistence in volatility is quite different between the two return series. This could be due to the two different sectors the stocks belong to and the investors belief of the riskiness not only of the two sectors but also of the two stocks. Table 4 provides values for the posterior median and the 95% credible intervals for the

	$n = 30$	$n = 50$	$n = 70$
c	7.7 (4.6, 13.3)	445.8 (146.2, 1088.9)	3822.7 (1697.7, 9858.5)
ζ	0.6 (0.3, 1.6)	0.7 (0.6,0.9)	1.0 (0.8, 1.3)

Table 4: Apple data: The posterior median and 95% credible interval for c and ζ for $n = 30$, $n = 50$, and $n = 70$.

concentration parameter c and ζ for $n = 30$, $n = 50$, and $n = 70$. The median value of ζ marginally increases across the three values of n, and the 95% credible intervals are narrower for the higher values of n. The median values for c increase as the value of n increases, and in this case they are in line with the values we had in our simulated example. The median value of c increases to 445.8 and 3822.7 for $n = 50$ and $n = 70$

respectively, hinting to a small variance for $F_\phi^{(n)}$. This implies that $F_\phi^{(n)}$ is close to a beta distribution, unlike the results with the HSBC returns series, which hinted that the distribution for ϕ may not be a beta.

1 week	2 weeks	8 weeks	1/2 year	1 year	2 years	5 years
0.24	0.40	0.65	0.78	0.85	0.90	0.96

Table 5: Apple data: The values of γ_κ with $n = 70$ for various lags

In Figure 5 we compare the posterior median volatility (together with 95% credible interval) for HSBC and Apple under our model with $n = 70$, to that of an SV model with a single AR(1) volatility process. The peaks and the day to day jumps in volatility are more evident under our model than the simple model. For HSBC we can identify two big jumps in volatility, one from September 2001 to November 2001(due to 9/11) and another from September 2008 to March 2009 (due to the start of the financial crisis following the collapse of the US subprime mortgage market). In the case of Apple there is one big surge in volatility around the end of September 2000 beginning of October 2000 which was due to the introduction of the OS X operating system and the iMac leading to a 30% boost of sales revenue. Since then Apple had been performing very well, with steady volatility, until September 2008 when sales revenue went up by 50% due to the introduction of the iPod touch and the iPhone 3G earlier that year.

The SV-ICA model expresses the volatility process in terms of sub-processes with different levels of dependence. It is difficult to interpret each individual sub-process and so we decompose into short term, medium term and long term components which aggregate some of these sub-processes according to their dependence. We define the components as:

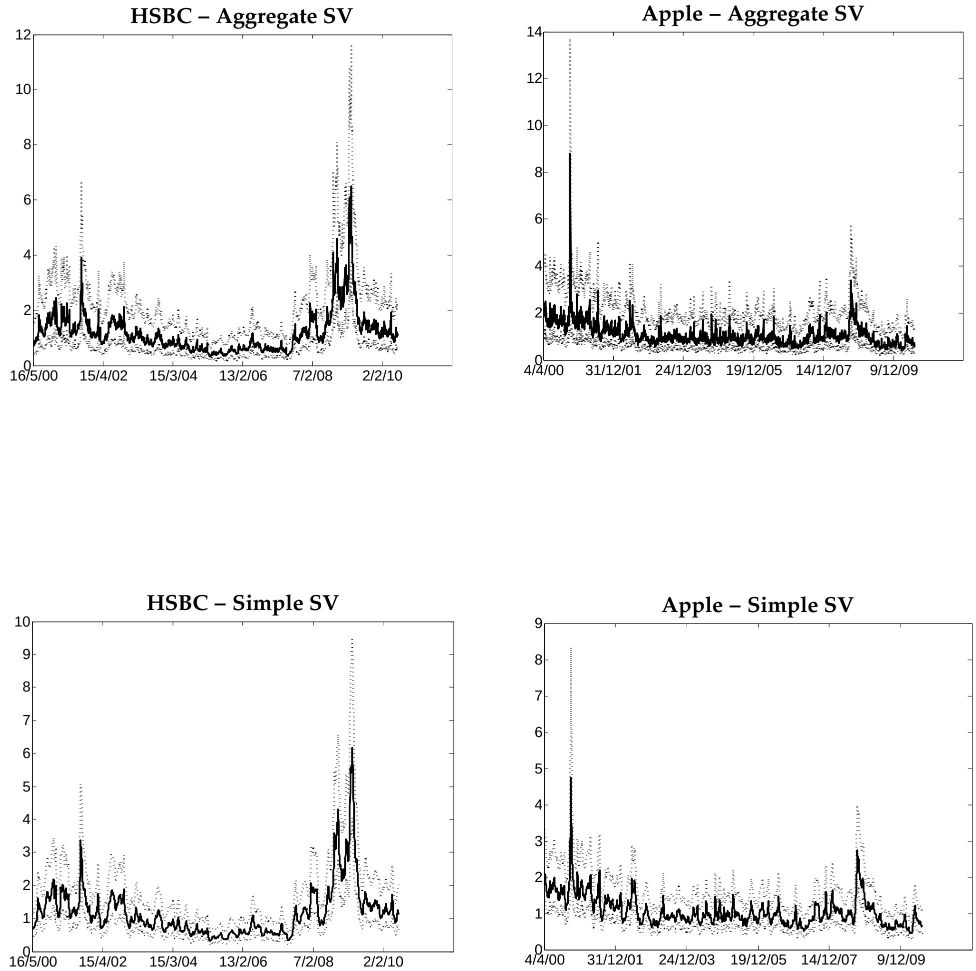

Figure 5: Posterior median volatility for HSBC and Apple (solid line) with 95% credible interval (dashed lines) for $n = 70$ - (top panel) for the simple SV model - (bottom panel)

$$h_t^{short} = \sum\nolimits_{\left\{i|\lambda_i<\epsilon^{1/126.5}\right\}} h_{t,i},$$

$$h_t^{med} = \sum\nolimits_{\left\{i|\epsilon^{1/126.5}<\lambda_i<\epsilon^{1/506}\right\}} h_{t,i},$$

$$h_t^{long} = \sum\nolimits_{\left\{i|\lambda_i>\epsilon^{1/506}\right\}} h_{t,i}.$$

This decomposition into components links to the quantity γ_κ defined in equation 15.

The short term component includes all processes whose dependence decays below ϵ by half a year, the medium term component includes all processes whose dependence decays below ϵ between half and two years, and the long term component includes all processes whose dependence decays below ϵ after two years.

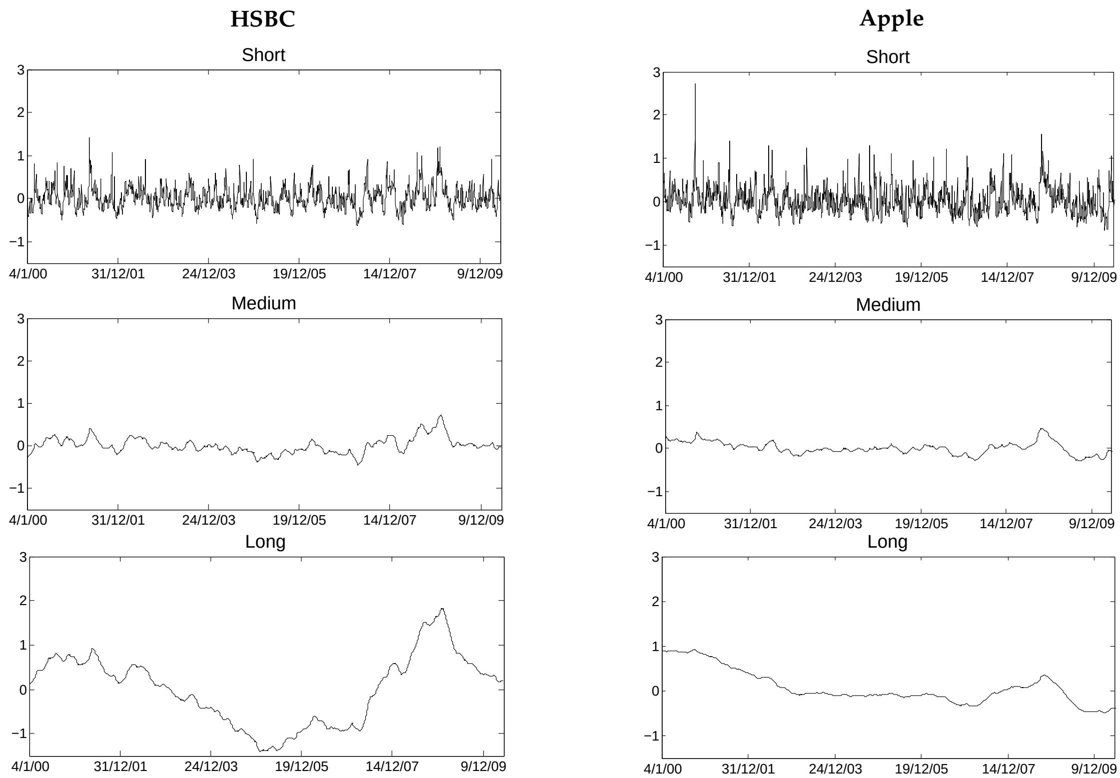

Figure 6: The posterior mean of the low, medium and high frequency components of the volatility process for HSBC and Apple.

In Figure 6, we decompose the aggregate volatility process of Apple Inc and HSBC plc, for $n = 70$, into short, medium and long term components. The two stocks show different evolutions of their long-term component. The long-term component of Apple has an overall decreasing trend, whereas there is no clear overall trend for HSBC. The effect of the financial crisis also has a very different effect on the volatilities with a much larger effect on HSBC compared to Apple, as we would expect. The unusual observation of Apple around September and October 2000 is included in the short-term volatility component with the volatility in that component leaping to around 2.5 and quickly decaying and with no effect on the long term component. This shows a robustness of the SV-ICA model to unusual observations.

4.3 Predictive performance

The predictive performance of our model is compared to a simple SV model with an AR(1) process for the log volatility and normal return distribution and a Bayesian semi-parametric model proposed by Jensen and Maheu (2010). They develop a stochastic volatility model with a Dirichlet process mixture model for the return distribution and an AR(1) process for the log volatility process. The comparison contrasts two ways of constructing a semiparametric stochastic volatility model. Our model nonparametrically models the dependence in the volatility process but retains a normal return distributions whereas Jensen and Maheu (2010) use a nonparametric return distribution with a parametric volatility process.

To access predictive performance we use the log predictive score (LPS) (Gneiting and Raftery, 2007) at different prediction horizons τ. In this case, the LPS is

$$\text{LPS}(\tau) = -\frac{1}{T-\tau-\lfloor T/2\rfloor+1}\sum_{i=\lfloor T/2\rfloor}^{T-\tau}\log p\left(y_i^{\tau}\,|\,y_1,\ldots,y_{i-1}\right)$$

where τ is a positive integer and $y_i^{\tau} = y_{i+\tau} - y_i$ is the log return over τ days. The results are presented for time horizon up to 150 days. The LPS is a strictly proper scoring rule with smaller values indicating a model giving better predictions. Figure 7 displays the LPS as a function of the forecasting horizon. Our SV-ICA model dominates the model with a nonparametric return distribution at all time horizons for the Apple data and at longer time horizons for the HSBC data. In both cases, the difference become

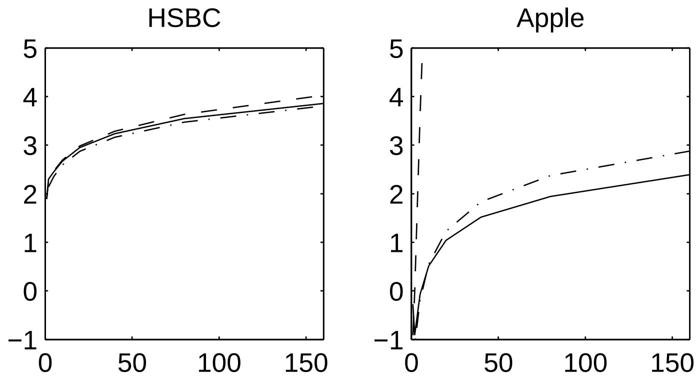

Figure 7: Log predictive score LPS(τ) as a function of forecasting horizon (τ). For the SV-ICA model with $n = 70$ (solid line), the simple SV model (dot-dashed line) and the Bayesian nonparametric model (dashed line) for HSBC and Apple.

larger at longer time horizons. The simple SV model performs well on the HSBC data (with a similar predictive performance to the other models) but is outperformed by the SV-ICA model in the Apple data. These findings suggest flexible modelling of the volatility dynamics results in better predictive performance when compared to flexibly modelling the return distribution.

5 Discussion

The SV-ICA model is described to flexibly model the dependence of the volatility process in financial time series. It is assumed that the volatility process is modelled using an ICA process which is formed as the limit of the aggregation of a finite number of AR(1) processes where the persistence parameter of each AR(1) process is independently drawn from a distribution F_ϕ. The distribution F_ϕ is given a Bayesian nonparametric prior. The infinite-dimensional prior for F_ϕ allows our model to adjust its complexity as more data is observed. The model can be interpreted as either providing a flexible prior for persistence or, alternatively, as modelling the inhomogeneity in the information flow driving the volatility by assuming that the effect of different information decays at different rates which can be modelled by an AR(1) process. Inference is made using a finite approximation to the Dirichlet process prior for F_ϕ with n elements. Convergence can be checked by looking at the posterior distribution for various values of n. An MCMC scheme to simulate from the posterior distribution of the finite approximation is proposed which uses adaptive MCMC ideas to provide an efficient sampler.

The method is illustrated on both simulated data, where the distribution used to generate the data can be accurately estimated, and two daily returns series (HSBC plc and Apple Inc.). The results in the real returns series show very different distributions for the persistence in volatility, with HSBC plc having much more mass on persistence which decayed more slowly than Apple Inc. This suggests the information related to the value of HSBC plc has much longer lasting effect.

The decomposition of the volatility process into multiple AR(1) processes allows us to better understand the dependence in the volatility process and conclude that a flexible model for persistence is more appropriate. This is so because the volatility of daily returns of stocks belonging to different sectors is affected by different flows of information. We plan in future work to consider the application of these methods to multiple returns series which will allow us to consider the amounts of information shared by different assets. More insight on the dependence between return series is crucial for investment, risk and portfolio managers.

References

J. Arteche. Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models. *Journal of Econometrics*, 119:131–154, 2004.

Y. F. Atchadé and J. S. Rosenthal. On Adaptive Markov Chain Monte Carlo Algorithms. *Bernoulli*, 11:815–828, 2005.

T. Bollerslev and H. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184, 1996.

F. J. Breidt, N. Grato, and P. De Lima. On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348, 1998.

F. Caron and A. Doucet. Sparse Bayesian nonparametric regression. In *Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*, 2008.

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

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

P. De Lima and N. Grato. Long-range dependence in the conditional variance of stock returns. *Economics Letters*, 45:281–285, 1994.

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

T. Ferguson. Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, 1:209–230, 1973.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

W. Fuller. *Introdiction to Time Series*. New York: John Wiley, 2nd edition, 1996.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction and estimation. *Journal of the American Statistical Association*, 102:359–378, 2007.

C. Granger. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.

J. E. Griffin. Inference in infinite superpositions of non-Gaussian Ornstein-Uhlenbeck processes with Bayesian nonparametric methods. *Journal of Financial Econometrics*, 9:519–549, 2011.

J. E. Griffin and P. J. Brown. Inference with Normal-Gamma prior distributions in regression problems. *Bayesian Analysis*, 5:171–188, 2010.

J. E. Griffin and M. F. J. Steel. Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics*, 123:121–152, 2004.

H. Haario, E. Saksman, and J. Tamminen. Compontentwise adaptation for high dimensional MCMC. *Computational Statistics*, 20:265–273, 2005.

A. Harvey. Long memory in stochastic volatility. In J. Knight and S. Satchell, editors, *Forecasting Volatility in Financial Markets*. Oxford: Butterworth-Heineman, 1998.

N. L. Hjort, C. Holmes, P. Müller, and S. Walker, editors. *Bayesian Nonparametrics*. Statistic and Probabilistic Mathematics. Cambridge University Press, 1st edition, April 2010.

C. Hurvich, E. Moulines, and P. Soulier. Estimating Long Memory in Volatility. *Econometrica*, 73(4):1283–1328, 2005.

H. Ishwaran and M. Zarepour. Markov Chain Monte Carlo in approximate Dirichlet and two-parameter process hierarchical models. *Biometrika*, 87:371–390, 2000.

H. Ishwaran and M. Zarepour. Exact and Approximate sum-Representations for the Dirichlet Process. *The Canadian Journal of Statistics*, 30:269–283, 2002.

M. J. Jensen and J. M. Maheu. Bayesian semiparametric volatility modeling. *Journal of Econometrics*, 157(2):306–316, 2010.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood Inference and Comparison with ARCH models. *Review of Economic Studies*, 65:361–93, 1998.

K. Latuszynski, G. O. Roberts, and J. S. Rosenthal. Adaptive Gibbs samplers and related MCMC methods. *The Annals of Applied Probability*, 23(1):66–98, 2013.

J. Nakajima and Y. Omori. Leverage, heavy-tails and correlated jumps in stochastic volatility models. *Computational Statistics & Data Analysis*, 53:2335–2353, 2009.

R. Neal. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2000.

Y. Omori, S. Chib, N. Shephard, and J. Nakajima. Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, 140:425–449, 2007.

S. Richardson and P. Green. Modelling heterogeneity with and without the Dirichlet process. *Scandinavian Journal of Statistics*, 28:355–377, 2001.

G. O. Roberts and J. S. Rosenthal. Examples of Adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18:349–367, 2009.

P. Robinson. Statistical inference for a randon coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5:163–168, 1978.

D. Surgailis and M. Viano. Long memory properties and covariance structure of the EGARCH model. *ESAIM*, 6:311–329, 2002.

S. Taylor. *Modeling Financial Time Series*. Wiley, Chichester, 1986.

S. J. Taylor. *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, 2005.

R. S. Tsay. *Analysis of Financial Time Series*. Probability and Statistics. Wiley, 2nd edition, 2005.

P. Zaffaroni. Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120:75–102, 2004.