

## Research Article

# Bayesian Non-Parametric Mixtures of GARCH(1,1) Models

**John W. Lau and Ed Cripps**

*School of Mathematics and Statistics, The University of Western Australia, Perth, Australia*

Correspondence should be addressed to John W. Lau, john.lau@uwa.edu.au

Received 2 March 2012; Revised 16 May 2012; Accepted 18 May 2012

Academic Editor: Ori Rosen

Copyright © 2012 J. W. Lau and E. Cripps. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional GARCH models describe volatility levels that evolve smoothly over time, generated by a single GARCH regime. However, nonstationary time series data may exhibit abrupt changes in volatility, suggesting changes in the underlying GARCH regimes. Further, the number and times of regime changes are not always obvious. This article outlines a nonparametric mixture of GARCH models that is able to estimate the number and time of volatility regime changes by mixing over the Poisson-Kingman process. The process is a generalisation of the Dirichlet process typically used in nonparametric models for time-dependent data provides a richer clustering structure, and its application to time series data is novel. Inference is Bayesian, and a Markov chain Monte Carlo algorithm to explore the posterior distribution is described. The methodology is illustrated on the Standard and Poor's 500 financial index.

## 1. Introduction

Generalised autoregressive conditional heteroscedastic (GARCH) models estimate time-varying fluctuations around mean levels of a time series known as the volatility of a time series [1, 2]. The standard GARCH model specifies the volatility persistence at time  $t$  as a linear combination of previous volatilities and squared residual terms. The persistence is assumed constant for all  $t$  resulting in smooth transitions of volatility levels. However, many nonstationary time series exhibit abrupt changes in volatility suggesting fluctuating levels of volatility persistence. In this case the GARCH parameters undergo regime changes over time. If the maximum number of potential regimes is known Markov-switching GARCH, models are an appealing option [3–8]. However, often the number of volatility regimes is not known and can be difficult to preselect. In this case, Bayesian nonparametric mixture models are attractive because they allow the data to determine the number of volatility regimes or mixture components. For example, recently nonparametric mixing over the Dirichlet process

has been described by Lau and So [9] for GARCH(1,1) models, Jensen and Maheu [10] for stochastic volatility models and Griffin [11] for Ornstein-Uhlenbeck Processes.

The Dirichlet process [12] is most widely applied in nonparametric mixture models and within a hierarchical framework is introduced in Lo [13] for independent data. Lau and So [9] extend the work of Lo [13] to time-dependent data where time-varying GARCH parameters are mixed over the Dirichlet process. The additional flexibility of this approach allows a range of GARCH regimes, from all observations generated by a single GARCH model to each observation in the series generated by a unique set of GARCH parameters. Lau and So [9] conclude with a discussion on the possibility of extending their method to alternative random probability measures that provide greater clustering flexibility than the Dirichlet process. We continue this discussion by outlining a novel method for a class of GARCH mixture models mixed over the Poisson-Kingman process [14, 15] derived from the stable subordinator (known henceforth as PKSS). To illustrate the richer clustering mechanisms of the PKSS process, we describe three of its special cases—the Dirichlet process [12], the Poisson-Dirichlet (PD) process [14, 16], and the Normalized Generalized Gamma (NGG) process [17, 18].

Theoretical developments and recent applications of the PKSS process are discussed in Lijoi et al. [19]. However, we note that the PD and the NGG processes have yet to be developed for volatility estimation, or indeed time series applications in general, and in this sense the work in this paper is novel. Although the Dirichlet process has now been used extensively in applications, the implementation of more general nonparametric mixture models for applied work is not always obvious. We therefore describe three Markov chain Monte Carlo (MCMC) algorithms. First, we develop a weighted Chinese restaurant Gibbs type process for partition sampling to explore the posterior distribution. The basis of this algorithm is developed for time-dependent data in Lau and So [9], and we extend it to allow for the PKSS process. We also note recent developments for the sampling of Bayesian nonparametric models in Walker [20] and Papaspiliopoulos and Roberts [21] and describe how these algorithms can be constructed to estimate our model.

The methodology is illustrated through volatility and predictive density estimation of a GARCH(1,1) model applied to the Standard and Poor's 500 financial index from 2003 to 2009. Results are compared between a no-mixture model and nonparametric mixtures over the Dirichlet, PD, and NGG processes. Under the criterion of marginal likelihood the NGG process performs the best. Also, the PD and NGG process outperforms the previously studied Dirichlet process which in turn outperforms the no-mixture model. The results suggest that alternatives to the Dirichlet process should be considered for applications of nonparametric mixture models to time-dependent data.

The paper proceeds as follows. Section 2 presents a Bayesian mixture of GARCH models over an unknown mixing distribution, outlines a convenient Bayesian estimator based on quadratic loss, and describes some of the time series properties of our model. Section 3 discusses the class of random probability measures we consider as the mixing distributions and details the clustering mechanisms associated with the three special cases mentioned above via the *Pólya Urn* representation and the consequences for the posterior distribution of the partitions resulting from the PKSS process. Our MCMC algorithm for sampling the posterior distribution is presented in Section 4, and the alternative MCMC algorithms of Walker [20] and Papaspiliopoulos and Roberts [21] are presented in Section 5. Section 6 describes the application, and Section 7 concludes the paper.

## 2. The Mixture Model

Let  $Y_t$  be the observed variable at time  $t$ ,  $\mathbf{Y}_{t-1} = \{Y_1, \dots, Y_{t-1}\}$  be the observations from time 1 to time  $t - 1$  and  $Y_t = 0$  for  $t \leq 0$ . The GARCH(1,1) model specifies

$$\begin{aligned} Y_t | \mathbf{Y}_{t-1}, (\nu, \chi, \psi) &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \nu + \chi Y_{t-1}^2 + \psi \sigma_{t-1}^2, \end{aligned} \quad (2.1)$$

where  $\nu > 0$ ,  $\chi \geq 0$ ,  $\psi \geq 0$  and  $\sigma_t^2 = 0$  for  $t \leq 0$ . In (2.1) the GARCH parameters  $(\nu, \chi, \psi)$  are not time varying implying that volatility persistence is constant over time with smooth transitions of volatility levels. To allow abrupt changes to volatilities, we extend (2.1) by writing  $Z_t = \{\nu_t, \chi_t, \psi_t\}$ ,  $\nu_t > 0$ ,  $\chi_t \geq 0$ ,  $\psi_t \geq 0$  for  $t = 1, \dots, n$  and  $\mathbf{Z}_t = \{Z_1, \dots, Z_t\}$  as joint latent variables from time 1 to time  $t$ ; that is, the model is now a dynamic GARCH model with each observation potentially generated by its own set of GARCH parameters as follows:

$$\begin{aligned} Y_t | \mathbf{Y}_{t-1}, \mathbf{Z}_t &\sim N(0, \sigma_t^2), \\ \sigma_t^2 &= \nu_t + \chi_t Y_{t-1}^2 + \psi_t \sigma_{t-1}^2. \end{aligned} \quad (2.2)$$

Note that in model (2.2) the data controls the maximum potential number of GARCH regimes, the sample size  $n$ . In contrast, finite switching models preallocate a maximum number of regimes typically much smaller than the number of observations. As the potential number of regimes gets larger, estimation of the associated transition probabilities and GARCH parameters in finite switching models become prohibitive. However, assuming that the latent variables  $\mathbf{Z}_n = \{Z_1, \dots, Z_n\}$  are independent of each other and completing the hierarchy by modelling the GARCH parameters contained in  $\mathbf{Z}_n$  with an unknown mixing distribution,  $G$ , with law  $\rho$  the model becomes manageable, that is,

$$\begin{aligned} Z_t | G &\stackrel{\text{iid}}{\sim} G(d\nu, d\chi, d\psi), \\ G &\sim \rho \end{aligned} \quad (2.3)$$

with  $\mathbf{Z}_n$  and the mixing distribution,  $G$ , parameters that we may estimate. Depending on the posterior distribution of the clustering structure associated with the mixing distribution, the results may suggest that  $Z_t = Z$  for  $t = 1, 2, \dots, n$  (a single regime GARCH model) up to a unique  $Z_t$  for each  $t$  indicating a separate GARCH regime for each time point. This illustrates the flexibility of the model.

We write  $s(\mathbf{Z}_n, G)$  as a positive integrable function of the latent variables,  $\mathbf{Z}_n$ , and the mixing distribution,  $G$ , to represent various quantities that may be of interest for inference. Under quadratic loss the Bayesian estimator is the posterior expectation  $E[s(\mathbf{Z}_n, G) | \mathbf{Y}_n]$ . For our model this is an appealing estimator because it does not require the posterior of  $G$  but only the posterior distribution of the sequence  $\{Z_1, \dots, Z_n\}$ , that is,

$$E[s(\mathbf{Z}_n, G) | \mathbf{Y}_n] = \int_{\mathcal{Z}^n} h(\mathbf{Y}_n, \mathbf{Z}_n) \pi(d\mathbf{Z}_n | \mathbf{Y}_n), \quad (2.4)$$

because

$$h(\mathbf{Y}_n, \mathbf{Z}_n) = \int_{\mathcal{G}} s(\mathbf{Z}_n, G) \pi(dG | \mathbf{Y}_n, \mathbf{Z}_n), \quad (2.5)$$

where  $\pi(dG | \mathbf{Y}_n, \mathbf{Z}_n)$  represents the posterior law of the random probability measure  $G$ , and  $\pi(d\mathbf{Z}_n | \mathbf{Y}_n)$  is the posterior distribution of the sequence  $\mathbf{Z}_n = \{Z_1, \dots, Z_n\}$ .

In Lau and So [9] the unknown mixing distribution for the GARCH parameters,  $G$ , is taken to be the Dirichlet process. This paper combines the theoretical ground work of Lijoi et al. [19, 22] with Lau and So [9] by allowing  $G$  to be the PKSS process. The result is a non-parametric GARCH model which contains (among others) the Dirichlet process typically used in time series as well as the PD and NGG processes as special cases.

Understanding conditions for stationarity of a time series model is fundamental for statistical inference. Since our model is specified with zero mean over time, we provide a necessary and sufficient condition for the existence of a secondary order stationary solution for the infinite mixture of GARCH(1,1) models. The derivation closely follows Embrechts et al. [23] and Zhang et al. [24], and we state the conditions without giving proof. By letting  $\epsilon_t$  be a standard normal random variable and replacing  $Y_{t-1}^2$  by  $\sigma_{t-1}^2 \epsilon_{t-1}^2$  then, conditioned on  $Y_i$  for  $i = 1, \dots, t-1$  and  $\{\nu_i, \chi_i, \psi_i\}$  for  $i = 1, \dots, t$ ,  $\sigma_t^2$  in (2.2) becomes

$$\sigma_t^2 = \nu_t + \chi_t Y_{t-1}^2 + \psi_t \sigma_{t-1}^2 = \nu_t + (\chi_t \epsilon_{t-1}^2 + \psi_t) \sigma_{t-1}^2. \quad (2.6)$$

Here (2.6) is well known to be an univariate stochastic difference equation expressed as

$$X_t = B_t + A_t X_{t-1}, \quad (2.7)$$

where  $X_t = \sigma_t^2$ ,  $A_t = \chi_t \epsilon_{t-1}^2 + \psi_t$ , and  $B_t = \nu_t$ . The stationarity of (2.7) implies the secondary order stationarity of (2.2), that is,  $X_t \xrightarrow{d} X$  as  $t \rightarrow \infty$  for some random variable  $X$ , and  $X$  satisfies  $X = B + AX$ , where the random variable pair  $(A_t, B_t) \xrightarrow{d} (A, B)$  as  $t \rightarrow \infty$  for some random variable pairs  $(A, B)$ . The stationary solution of (2.7) holds if  $E[\ln^+ |A_t|] < 0$  and  $E[\ln^+ |B_t|] < \infty$ , where  $\ln^+ |x| = \ln[\max\{x, 1\}]$  as given in Embrechts et al. [23, Section 8.4, pages 454–481] and Zhang et al. [24, Theorems 2 and 3], Vervaat [25], and Brandt [26]. So the conditions for the stationarity in our model are

$$\int_{(0, \infty)} \ln^+ |v| H_1(dv) < \infty, \quad \int_{(0, \infty)^2} E[\ln^+ |\chi \epsilon^2 + \psi|] H_2(d\chi, d\psi) < 0, \quad (2.8)$$

where the expectation for the second condition is applicable only to  $\epsilon$ , which is a standard normal random variable, and both  $H_1 = \int_{(0, \infty)^2} H(dv, d\chi, d\psi)$  and  $H_2 = \int_{(0, \infty)} H(dv, d\chi, d\psi)$  are marginal measures of  $H = E[G]$ , the mean measure of  $G$ .

Now consider the first two conditional moments from model (2.2). Obviously, the first conditional moment is zero, and the second conditional moment is identical to  $\sigma_t^2$  in (2.6). The distinguishing feature of model (2.2) is that parameters change over time and have the

distribution  $G$ . Considering  $\sigma_t^2$  as a scale of the model results in a scale mixture model over time. From the representation in (2.6),  $\sigma_t^2$  could be rewritten as

$$\sigma_t^2 = \nu_t + \sum_{j=1}^{t-1} \nu_{t-j} \prod_{i=t-j+1}^t (\chi_i e_{i-1}^2 + \psi_i). \quad (2.9)$$

The unconditional second moment could be derived according to this representation by marginalising over all the random variates. Also,  $\sigma_t^2$  in (2.9) can be viewed as a weighted sum of the random sequence  $\{\nu_t, \dots, \nu_1\}$ , and the random weights decay to zero at a polynomial rate, as long as the model is stationary. In fact, the rate could be irregular over time, and this is a substantial difference between the mixture of GARCH models and the traditional GARCH models.

Finally, one might be also interested in the connection between models such as (2.2) with parameters having the distribution (2.3) and those having the Markov-switching characteristic that result in Markov-switching GARCH models [6, 7]. Markov-switching GARCH models have a similar structure to (2.3), and we can replace (2.3) by

$$\begin{aligned} Z_t | (S_t = i) &= \{\nu_i, \chi_i, \psi_i\}, \\ S_t | S_{t-1} &\sim \eta_{ij} = P(S_t = i | S_{t-1} = j), \quad \text{for } i, j = 1, \dots, K, \end{aligned} \quad (2.10)$$

where  $S_t$  denotes the state variables, usually latent and unobserved. Marginalising the current state variable  $S_t$  in (2.10) yields the conditional distribution for  $Z_t$  given the previous state  $S_{t-1}$ ,

$$Z_t | S_{t-1} \sim \sum_{i=1}^K \eta_{ij} \delta_{(\nu_i, \chi_i, \psi_i)}. \quad (2.11)$$

So (2.11) could be a random probability measure but with a finite number of components and dependent on previous state  $S_{t-1}$ .

### 3. The Random Probability Measures

We now describe PKSS process and detail the Dirichlet, the PD, and NGG processes to illustrate how the more general PKSS process allows for richer clustering mechanisms. Let  $\mathcal{Z}$  be a complete and separable metric space and  $\mathcal{B}(\mathcal{Z})$  be the corresponding Borel  $\sigma$ -field. Let  $G \in \mathcal{G}$  be a probability measure on the space  $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$  where  $\mathcal{G}$  is the set of probability measures equipped with the suitable  $\sigma$ -field  $\mathcal{B}(\mathcal{G})$  and the corresponding probability measure  $\mathcal{D}$  (see chapter 2 of Ghosh and Ramamoorthi [27] for more details). The random probability measure  $G$  is sampled from the law  $\mathcal{D}$  and operates as the unknown mixing distribution of the GARCH parameters in (2.2).

All random probability measures within the class of PKSS processes feature an almost surely discrete probability measure represented as

$$G(A) = \sum_{i=1}^{\infty} W_i \delta_{Z_i}(A) \quad \text{for } A \in \mathcal{B}(\mathcal{Z}), \quad (3.1)$$

where  $\delta_{Z_i}$  denotes the dirac delta measure concentrated at  $Z_i$ , in which the sequence of the random variables  $\{Z_1, Z_2, \dots\}$  has a nonatomic probability measure  $H$  and the sequence of the random variables  $\{W_1, W_2, \dots\}$  sums to 1 [28]. Also, the mean measure of the process  $G$  with respect to  $\mathcal{P}$  is  $H$  as follows:

$$E[G(A)] = H(A) \quad \text{for } A \in \mathcal{B}(\mathcal{Z}). \quad (3.2)$$

A common characterization of (3.1) is the well-known *Pólya Urn* prediction distribution described in Pitman [28]. For the purposes of this paper the *Pólya Urn* warrants further discussion for two reasons. First, it is important for developing our MCMC algorithm to explore the posterior distribution discussed in Section 4. Second, it explicitly details how the PKSS process is a generalisation of the Dirichlet, PD, and NGG processes and how the different cluster tuning mechanisms operates.

Let  $\{Z_1, \dots, Z_r\}$  be a sequence with size  $r$  drawn from  $G$  where  $r$  is a positive integer, and let  $\mathbf{p}_r$  denote a partition of integers  $\{1, \dots, r\}$ . A partition  $\mathbf{p}_r = \{C_{r,1}, \dots, C_{r,N_r}\}$  of size  $N_r$  contains disjoint clusters  $C_{r,j}$  of size  $e_{r,j}$  indicated by the distinct values  $\{Z_1^*, \dots, Z_{N_r}^*\}$ . The *Pólya Urn* prediction distribution for the PKSS process can now be written as

$$\pi(dZ_{i+1} | \mathbf{Z}_i) = \frac{V_{i+1, N_i+1}}{V_{i, N_i}} H(dZ_{i+1}) + \frac{V_{i+1, N_i}}{V_{i, N_i}} \sum_{j=1}^{N_i} (e_{i,j} - \alpha) \delta_{Z_j^*}(dZ_{i+1}), \quad (3.3)$$

for  $i = 1, \dots, r-1$ ,  $\pi(dZ_1) = H(dZ_1)$ ,  $V_{1,1} = 1$  and

$$V_{i, N_i} = (i - N_i \alpha) V_{i+1, N_i} + V_{i+1, N_i+1}. \quad (3.4)$$

The *Pólya Urn* prediction distribution is that  $Z_{i+1}$  will take a new value from  $H$  with mass  $V_{i+1, N_i+1}/V_{i, N_i}$  and one of the existing values,  $\{Z_1^*, \dots, Z_{N_i}^*\}$ , with mass  $(i - N_i \alpha) V_{i+1, N_i} / V_{i, N_i}$ . This yields a joint prior distribution

$$\pi(dZ_1, dZ_2, \dots, dZ_n) = \pi(dZ_1) \prod_{i=1}^{n-1} \pi(dZ_{i+1} | Z_1, \dots, Z_i), \quad (3.5)$$

as the product of easily managed conditional densities useful for our MCMC scheme below.

The PKSS process can be represented as either the Dirichlet, PD, or NGG processes in (3.3) as the following.

(1) Taking  $0 \leq \alpha < 1$  and

$$V_{i, N_i} = \frac{\Gamma(i)}{\Gamma(i + \theta)} \prod_{j=1}^{N_i} (\theta + (j-1)\alpha), \quad (3.6)$$

for  $\theta > -\alpha$  results in the PD process.

(2) Setting that  $\alpha = 0$

$$V_{i,N_i} = \frac{\Gamma(i)}{\Gamma(i+\theta)} \theta^{N_i}, \quad (3.7)$$

and the PD process becomes the Dirichlet process.

(3) The NGG process takes  $0 \leq \alpha < 1$ , such that

$$V_{i,N_i} = \frac{\alpha^{N_i-1} e^{\beta} \sum_{k=0}^{i-1} \binom{i-1}{k} (-1)^k \beta^{k/\alpha} \Gamma\left(N_i - \frac{k}{\alpha}, \beta\right)}{\Gamma(i)}, \quad (3.8)$$

for  $\beta > 0$ .

In the above  $\Gamma(\cdot)$  is the complete Gamma function, and  $\Gamma(\cdot, \cdot)$  is the incomplete Gamma function. Examining of the predictive distribution (3.3), the ratios  $V_{i+1,N_{i+1}}/V_{i,N_i}$  and  $V_{i+1,N_i}/V_{i,N_i}$  indicate the difference between the Dirichlet process and the other processes. Substituting the values of  $V_{i,N_i}$  into the allocation mass for each process reveals that the ratios do not depend on the number of existing clusters,  $N_i$ . Rather, the Dirichlet process assigns probability to a new value independent of the number of existing clusters, and the rate of increment of partition size is a constant. In contrast, the PD and NGG processes assign probability to a new value dependent on the number of existing clusters. The comparison of these three special cases illustrates the richer clustering mechanisms of the PKSS process over the Dirichlet process. Furthermore, the PKSS process contains many other random measures, and these measures would be of interest for their clustering behaviors in further investigation.

Turning to the distribution of partitions, Pitman [28] shows that the joint prior distribution of the sequence  $\{Z_1, \dots, Z_n\}$  is

$$\pi(dZ_1, \dots, dZ_r) = V_{r,N_r} \prod_{j=1}^{N_r} \frac{\Gamma(e_{r,j} - \alpha)}{\Gamma(1 - \alpha)} H(dZ_j^*). \quad (3.9)$$

Notice that the joint distribution is dependent on the partition  $\mathbf{p}_r$  of  $r$  integers  $\{1, \dots, r\}$ , and we can decompose (3.9) into  $\pi(Z_1^*, \dots, Z_{N_r}^*, \mathbf{p}_r) = \pi(dZ_1^*, \dots, dZ_{N_r}^* | \mathbf{p}_r) \pi(\mathbf{p}_r)$ . The distribution of the partition,  $\pi(\mathbf{p}_r)$ , is

$$\pi(\mathbf{p}_r) = V_{r,N_r} \prod_{j=1}^{N_r} \frac{\Gamma(e_{r,j} - \alpha)}{\Gamma(1 - \alpha)} \quad (3.10)$$

and is known as the *Exchangeable Partition Probability Function*. For many nonparametric models, this representation also helps MCMC construction by partitioning the posterior distribution in the form of *Exchangeable Partition Probability Function*. To do so it is necessary to obtain the posterior distribution of the partition  $\pi(\mathbf{p}_n | \mathbf{Y}_n)$  analytically. Then we could generate  $\mathbf{p}_n$  and approximate the posterior expectation. However, this is not possible in general.

So we consider the joint distribution of  $\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}$  instead. We write the posterior expectation of  $s(\mathbf{Z}_n, G)$  as a marginalization over the joint posterior distribution of  $\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}$  by

$$E[s(\mathbf{Z}_n, G) | \mathbf{Y}_n] = \sum_{\mathbf{p}_n} \underbrace{\int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}}}_{N_n} h(\mathbf{Y}_n, \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n). \quad (3.11)$$

Here the joint posterior distribution of  $(\mathbf{Z}_{N_n}^*, \mathbf{p}_n)$  is given by

$$\pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n) = \frac{\prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\})) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n)}{\sum_{\mathbf{p}_n} \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\})) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n)}, \quad (3.12)$$

where  $\phi(x | a, b)$  represents a normal density with mean  $a$  and variance  $b$  evaluated at  $x$ . The variance  $\sigma_t^2(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\})$  is identical to  $\sigma_t^2$  and emphasizes that  $\sigma_t^2$  is a function of  $\{Z_1, Z_2, \dots, Z_t\}$  represented by  $(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\})$ . This representation leads to the development of the MCMC algorithm in the next section. For the sake of simplicity, we prefer the following expression for the variance:

$$\sigma_t^2(\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) = \sigma_t^2(\mathbf{Z}_n) = \sigma_t^2(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\}) = \sigma_t^2(\mathbf{Z}_t) = \sigma_t^2. \quad (3.13)$$

We emphasise that we can always express  $\mathbf{Z}_n$  by two elements, namely, a partition and distinct values. In this case  $\mathbf{p}_n$  is a partition of the integers  $\{1, \dots, n\}$  that are the indices of  $\mathbf{Z}_n$ , and  $\mathbf{Z}_{N_n}^* = \{Z_1^*, \dots, Z_{N_n}^*\}$  represents the distinct values of  $\mathbf{Z}_n$ . The partition  $\mathbf{p}_n$  locates the distinct values from  $\mathbf{Z}_n$  to  $\mathbf{Z}_{N_n}^*$  or vice versa. As a result, we have the following equivalent representations:

$$\mathbf{Z}_n = \{Z_1, \dots, Z_n\} = \{Z_1^*, \dots, Z_{N_n}^*\} = \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}. \quad (3.14)$$

In time series analysis, we usually consider the first  $t$  items,  $\mathbf{Z}_t = \{Z_1, \dots, Z_t\}$ , the corresponding partition,  $\mathbf{p}_t$ , and distinct values,  $\mathbf{Z}_{N_t}^*$ , such that

$$\mathbf{Z}_t = \{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\}. \quad (3.15)$$

Here  $\mathbf{Z}_t$  contains the first  $t$  elements of  $\mathbf{Z}_n = \{Z_1, \dots, Z_n\}$ , and adding  $\{t+1, \dots, n\}$  to  $\mathbf{p}_t$  would yield  $\mathbf{p}_n$  according to  $\mathbf{Z}_{N_t}^*$  and the distinct values of  $\{Z_{t+1}, \dots, Z_n\}$ . Combining  $\mathbf{Z}_{N_t}^*$  and the distinct values of  $\{Z_{t+1}, \dots, Z_n\}$  gives  $\mathbf{Z}_{N_n}^*$  providing the connection between  $\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\}$  and  $\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}$ . To simplify the likelihood expression and the sampling algorithm, we replace  $\sigma_t^2(\mathbf{Z}_t) = \sigma_t^2(\{\mathbf{Z}_{N_t}^*, \mathbf{p}_t\})$  by  $\sigma_t^2(\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\})$  since the subscript  $t$  in  $\sigma_t^2$  already tells us that the first  $t$  items of  $\mathbf{Z}_n$  are considered. We then have a more accessible representation of the likelihood function as

$$L(\mathbf{Y}_n | \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) = \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\})), \quad (3.16)$$



and (3.12) becomes

$$\pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n) = \frac{L(\mathbf{Y}_n | \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n)}{\sum_{\mathbf{p}_n} \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n)}. \quad (3.17)$$

We are now equipped to describe the MCMC algorithm.

#### 4. The Algorithm for the Partitions and Distinct Values Sampling

Our Markov chain Monte Carlo (MCMC) sampling procedure generates distinct values and partitions alternatively from the posterior distribution,  $\pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n)$ . For  $S$  iterations our MCMC algorithm is

- (1) Initialise  $\mathbf{Z}_{N_n}^* = (\mathbf{Z}_{N_n}^*)^{[0]}$ .  
For  $s = 1, 2, \dots, S$ .
- (2) Generate  $\mathbf{p}_n^{[s]}$  from  $\pi(\mathbf{p}_n^{[s]} | (\mathbf{Z}_{N_n}^*)^{[s-1]}, \mathbf{Y}_n)$ .
- (3) Generate  $(\mathbf{Z}_{N_n}^*)^{[s]}$  from  $\pi(d\mathbf{Z}_{N_n}^* | \mathbf{p}_n^{[s]}, \mathbf{Y}_n)$ .

End.

To obtain our estimates we use the weighted Chinese restaurant Gibbs type process introduced in Lau and So [9] for the time series models mixed over Dirichlet process. We have extended this scheme to allow for the more general PKSS process. In what follows, the extension from the Dirichlet to the PKSS lies in the weights of the *Pólya Urn* predictive distribution in (3.3).

The main idea of this algorithm is the “leave one out” principle that removes item  $t$  from the partition and then replaces it. This will give an update on both  $\mathbf{Z}_{N_n}^*$  and  $\mathbf{p}_n$ . This idea has been applied in sampling of partitions in many Bayesian nonparametric models of Dirichlet process (see [17] for an review). The strategy is a simple evaluation on the product of the likelihood function (3.16) and the *Pólya Urn* distribution (3.3) of  $Z_t$ , conditioned on the remaining parameters, yielding a joint updating distribution of  $\mathbf{Z}_{N_n}^*$  and  $\mathbf{p}_n$ . We now describe the distributions  $\pi(\mathbf{p}_n | \mathbf{Z}_{N_n}^*, \mathbf{Y}_n)$  and  $\pi(d\mathbf{Z}_{N_n}^* | \mathbf{p}_n, \mathbf{Y}_n)$  used in the sampling scheme.

Define  $\mathbf{p}_{n-t}$  to be the partition  $\mathbf{p}_n$  less item  $t$ . Then  $\mathbf{p}_{n-t} = \{C_{1,-t}, C_{2,-t}, \dots, C_{N_{n-t},-t}\}$  with corresponding distinct values given by  $\mathbf{Z}_{N_{n-t}}^* = \{Z_{1,-t}^*, Z_{2,-t}^*, \dots, Z_{N_{n-t},-t}^*\}$ . To generate from  $\pi(\mathbf{p}_n | \mathbf{Z}_{N_n}^*, \mathbf{Y}_n)$  for each  $t = 1, \dots, n$ , the item  $t$  is assigned either to a new cluster  $C_{N_{n-t}+1,-t}$ ; that is, empty before  $t$  is assigned with probability

$$\frac{\pi(\tilde{\mathbf{p}}_{N_{n-t}+1}) \times \int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_{N_{n-t}+1}^*, \tilde{\mathbf{p}}_{N_{n-t}+1}\}) H(dZ_{N_{n-t}+1,-t})}{\pi(\tilde{\mathbf{p}}_{N_{n-t}+1}) \times \int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_{N_{n-t}+1}^*, \tilde{\mathbf{p}}_{N_{n-t}+1}\}) H(dZ_{N_{n-t}+1,-t}) + \sum_{j=1}^{N_{n-t}} \pi(\tilde{\mathbf{p}}_j) \times L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_j^*, \tilde{\mathbf{p}}_j\})} \quad (4.1)$$

or to an existing cluster,  $C_{j,-t}$  for  $j = 1, \dots, N_{n,-t}$  with probability

$$\frac{\pi(\tilde{\mathbf{p}}_j) \times L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_j^*, \tilde{\mathbf{p}}_j\})}{\pi(\tilde{\mathbf{p}}_{N_{n,-t}+1}) \times \int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_{N_{n,-t}+1}^*, \tilde{\mathbf{p}}_{N_{n,-t}+1}\}) H(dZ_{N_{n,-t}+1,-t}) + \sum_{j=1}^{N_{n,-t}} \pi(\tilde{\mathbf{p}}_j) \times L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_j^*, \tilde{\mathbf{p}}_j\})}, \quad (4.2)$$

where

$$\begin{aligned} \tilde{\mathbf{p}}_j &= \mathbf{p}_{n,-t} \cup \{t \in C_{j,-t}\}, \\ \tilde{\mathbf{Z}}_j^* &= \{\mathbf{Z}_{N_{n,-t}-t}^*, Z_t = Z_{j,-t}^*\}, \end{aligned} \quad (4.3)$$

for  $j = 1, \dots, N_{n,-t} + 1$ . In addition, if  $j = N_{n,-t} + 1$  and a new cluster is selected, a sample of  $Z_{N_{n,-t}+1,-t}^*$  is drawn from

$$\frac{L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_{N_{n,-t}+1}^*, \tilde{\mathbf{p}}_{N_{n,-t}+1}\}) H(dZ_{N_{n,-t}+1,-t}^*)}{\int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\tilde{\mathbf{Z}}_{N_{n,-t}+1}^*, \tilde{\mathbf{p}}_{N_{n,-t}+1}\}) H(dZ_{N_{n,-t}+1,-t}^*)} \quad (4.4)$$

for the next iteration.

To generate from  $\pi(\mathbf{Z}_{N_n}^* | \mathbf{p}_n, \mathbf{Y}_n)$  for  $j = 1, \dots, N_n$  generate  $Z_j^*$  given  $\{Z_1^*, \dots, Z_{N_n}^*\} \setminus \{Z_j^*\}$ ,  $\mathbf{p}_n$ , and  $\mathbf{Y}_n$  from the conditional distribution

$$\pi(Z_j^* | \mathbf{Z}_{N_n}^* \setminus \{Z_j^*\}, \mathbf{p}_n, \mathbf{Y}_n) = \frac{L(\mathbf{Y}_n | \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) H(dZ_j^*)}{\int_{\mathcal{Z}} L(\mathbf{Y}_n | \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}) H(dZ_j^*)}. \quad (4.5)$$

This step uses the standard Metropolis-Hastings algorithm to draw the posterior samples  $Z_j^*$ . Precisely, (4.5) is given by

$$\pi(dZ_j^* | \mathbf{Z}_{N_n}^* \setminus \{Z_j^*\}, \mathbf{p}_n, \mathbf{Y}_n) = \frac{\prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\})) H(dZ_j^*)}{\int_{\mathcal{Z}} \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\})) H(dZ_j^*)}. \quad (4.6)$$

In (4.6) all elements in the sequence  $\{Z_1^*, \dots, Z_{N_n}^*\}$ , conditional on  $\mathbf{p}_n$  and  $\mathbf{Y}_n$ , are no longer independent, and they require to be sampled individually, and conditional the remaining elements.

We note that the special case of the above algorithm can be found for independent data of the normal mixture models from West et al. [29] (see also [17]). Taking  $\chi = 0$  and  $\psi = 0$  in

(2.2) yields  $Z_t = \nu_t = \sigma_t^2$  for  $i = 1, \dots, n$  with distribution  $G$ . Let  $G$  be a Dirichlet process with parameter  $\theta H$ . Then (4.1) and (4.2) become

$$\frac{\theta \times \int_{\mathcal{Z}} \phi(Y_t | 0, Z) H(dZ)}{\theta \times \int_{\mathcal{Z}} \phi(Y_t | 0, Z) H(dZ) + \sum_{j=1}^{N_{n-t}} e_{j,-t} \times \phi(Y_t | 0, Z_{j,-t}^*)}, \quad (4.7)$$

$$\frac{e_{j,-t} \times \phi(Y_t | 0, Z_{j,-t}^*)}{\theta \times \int_{\mathcal{Z}} \phi(Y_t | 0, Z) H(dZ) + \sum_{j=1}^{N_{n-t}} e_{j,-t} \times \phi(Y_t | 0, Z_{j,-t}^*)}.$$

Furthermore, the joint distribution of  $\{Z_1, \dots, Z_{N_n}\}$  conditional on  $\mathbf{p}_n$  and  $\mathbf{Y}_n$  is given by

$$\prod_{j=1}^{N_n} \prod_{t \in C_j} \phi(Y_t | 0, Z_j^*) H(dZ_j^*). \quad (4.8)$$

In this case  $\{Z_1, \dots, Z_{N_n}\}$  are independent in both the prior,  $\pi(\mathbf{Z}_{N_n}^* | \mathbf{p}_n)$ , and the posterior,  $\pi(\mathbf{Z}_{N_n}^* | \mathbf{p}_n, \mathbf{Y}_n)$ . However, it is not true in the more general dynamic GARCH model we consider.

Usually, the parameters of interest are both the sequence  $\{\sigma_1, \dots, \sigma_n\}$  and the predictive density  $\int_{\mathcal{Z}} k(Y_{n+1} | \mathbf{Y}_n, Z_n, Z_{n+1}) G(dZ_{n+1})$ . These two sets of parameters are functions of  $\mathbf{Z}_n$  under the mixture of GARCH(1,1) model, and the Bayesian estimators are taken to be the expected means as outlined in Section 2. That is, writing the volatility as the vector  $\sigma_n(\mathbf{Z}_n) = \{\sigma_1(\mathbf{Z}_n), \dots, \sigma_n(\mathbf{Z}_n)\}$

$$E[\sigma_n(\mathbf{Z}_n) | \mathbf{Y}_n] = \sum_{\mathbf{p}_n} \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} \sigma_n(\mathbf{Z}_{N_n}^*, \mathbf{p}_n) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n),$$

$$E\left[\int_{\mathcal{Z}} k(Y_{n+1} | \mathbf{Y}_n, Z_{n+1}, \mathbf{Z}_n) G(dZ_{n+1}) | \mathbf{Y}_n\right] = \sum_{\mathbf{p}_n} \int_{\mathcal{Z}} \cdots \int_{\mathcal{Z}} \int_{\mathcal{Z}} k(Y_{n+1} | \mathbf{Y}_n, Z_{n+1}, \mathbf{Z}_{N_n}^*, \mathbf{p}_n)$$

$$\times \pi(dZ_{n+1} | \mathbf{Z}_{N_n}^*, \mathbf{p}_n) \pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n), \quad (4.9)$$

where  $\pi(d\mathbf{Z}_{N_n}^*, \mathbf{p}_n | \mathbf{Y}_n)$  denotes the posterior distribution of  $\mathbf{Z}_n = \{\mathbf{Z}_{N_n}^*, \mathbf{p}_n\}$  and  $\pi(dZ_{n+1} | \mathbf{Z}_{N_n}^*, \mathbf{p}_n)$  represents the *Pólya Urn* predictive density of  $Z_{n+1}$  given  $\mathbf{Z}_n$ .

## 5. Alternative Algorithmic Estimation Procedures

We now outline how the algorithms of Walker [20] and Papaspiliopoulos and Roberts [21] may be applied to our GARCH(1,1) model mixed over the PKSS process. First, consider the approach in Walker [20]. Beginning with (3.1) the weights can be written as

$$W_i = \int_{(0, W_i)} du = W_i \int_{(0, \infty)} W_i^{-1} \mathbb{I}_{\{(0, W_i)\}}(u) du = W_i \int_{(0, \infty)} U(u | 0, W_i) du, \quad (5.1)$$

where  $\mathbb{I}_{\{(0, W_i)\}}(\mathbf{u})$  denotes the indicator function that  $\mathbb{I}_{\{(0, W_i)\}}(\mathbf{u}) = 1$  if  $0 < u < W_i$  and  $\mathbb{I}_{\{(0, W_i)\}}(\mathbf{u}) = 0$  otherwise, and  $U(u | 0, W_i)$  represents the uniform density on the interval  $(0, W_i)$ . Then, substituting (5.1), but without the integral over  $U$ , into (3.1), we obtain the joint measure

$$G(dz, du) = \sum_{i=1}^{\infty} W_i U(u | 0, W_i) du \delta_{Z_i}(dz). \quad (5.2)$$

Furthermore, we can take the classification variables  $\{\delta_1, \dots, \delta_n\}$  to indicate the points  $\{Z_{\delta_1}, \dots, Z_{\delta_n}\}$  taken from the measure. The classification variables  $\{\delta_1, \dots, \delta_n\}$  take values from the integers  $\{1, 2, \dots\}$  and assign a configuration to model (2.2) so that the expression of the likelihood is simpler without the product of sums. So, combining (5.2) with model (2.2) yields

$$L(\mathbf{Y}_n, \mathbf{u}_n, \boldsymbol{\delta}_n | \mathbf{Z}, \mathbf{W}) = \prod_{t=1}^n W_{\delta_t} U(u_t | 0, W_{\delta_t}) \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t})), \quad (5.3)$$

where  $\boldsymbol{\delta}_n = \{\delta_1, \dots, \delta_n\}$  and  $\mathbf{Z}_{\delta_t} = \{Z_{\delta_1}, \dots, Z_{\delta_t}\}$  for  $t = 1, \dots, n$ . Here the random jumps that build up the random measure  $G$  in (5.2) can be reexpressed as  $W_1 \stackrel{d}{=} V_1$  and  $W_i \stackrel{d}{=} V_i \prod_{j=1}^{i-1} (1 - V_j)$  for  $j = 2, 3, \dots$ . This is called the stick-breaking representation. Unfortunately, up to now, this representation includes only the Poisson-Dirichlet process  $(\alpha, \theta)$  where  $V_i$ s are Beta( $1 - \alpha, \theta + i\alpha$ ) random variables. Further development will be required to fully utilise the approach of Walker [20] for the PKSS process in general.

The likelihood (5.3) can be written as

$$L(\mathbf{Y}_n, \mathbf{u}_n, \boldsymbol{\delta}_n | \mathbf{Z}, \mathbf{V}) = \prod_{t=1}^n W_{\delta_t}(\mathbf{V}) U(u_t | 0, W_{\delta_t}(\mathbf{V})) \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t})), \quad (5.4)$$

and MCMC algorithms for sampling  $\mathbf{u}_n$ ,  $\boldsymbol{\delta}_n$ , and  $\mathbf{V}$  are straightforward and already included in Walker ([20], Section 3). To complete the algorithm for our model it requires sampling from  $\mathbf{Z}$ . This can be achieved by sampling from  $Z_j$  for all  $\{j : \delta_t = j\}$  from

$$\frac{\prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t})) H(dZ_j)}{\int_{\mathcal{Z}} \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t})) H(dZ_j)}, \quad (5.5)$$

otherwise sample  $Z_j$  from  $H(dZ)$  if there is no  $j$  equal to  $\delta_t$ . Notice that there are infinite  $Z_j$ s contained in  $\mathbf{Z}$ , but it is only required to sample at most  $n$  of them. The number of sampled  $Z_j$ s varies over iterations (see also ([20], Section 3) for details).

Papaspiliopoulos and Roberts [21] suggest an approach similar to Walker [20]. Consider the classification variables  $\{\delta_1, \dots, \delta_n\}$  and the stick-breaking representation of  $\{W_1, W_2, \dots\}$  contributed by  $\{V_1, V_2, \dots\}$  defined above. Then the likelihood is immediately given by

$$L(\mathbf{Y}_n, \boldsymbol{\delta}_n | \mathbf{Z}, \mathbf{V}) = \prod_{t=1}^n W_{\delta_t}(\mathbf{V}) \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t})). \quad (5.6)$$

The most challenging task is the reallocation of  $n$  observations over the infinite components in (3.1), equivalent to sampling the classification variables  $\{\delta_1, \dots, \delta_n\}$  over the MCMC iterations. Here we will briefly discuss this task when it involves the variance  $\sigma_t^2(\mathbf{Z}_{\delta_t})$  for our model (see also Papaspiliopoulos and Roberts [21, Section 3]). Let

$$\boldsymbol{\delta}_n(i, j) = \{\delta_1, \dots, \delta_{i-1}, j, \delta_{i+1}, \dots, \delta_n\} \quad (5.7)$$

be the vector produced from  $\boldsymbol{\delta}_n$  by substituting the  $i$ th element by  $j$ . This is a proposed move from  $\boldsymbol{\delta}_n$  to  $\boldsymbol{\delta}_n(i, j)$  where  $j = 1, 2, \dots$ . Notice that it is not possible to consider an infinite number of  $Z_j$ s directly since we only have a finite number. Instead, we can employ the Metropolis-Hastings sampler that considers a proposal probability mass function which requires only a finite number of  $Z_j$ s. The probabilities for the proposed moves are given by

$$q_n(i, j) = \frac{W_j}{C(\boldsymbol{\delta}_n)} \times \begin{cases} \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t}))|_{\delta_t=j}, & \text{for } j = 1, \dots, \max\{\delta_1, \dots, \delta_n\} \\ M(\boldsymbol{\delta}_n), & \text{for } j > \max\{\delta_1, \dots, \delta_n\}, \end{cases} \quad (5.8)$$

where  $M(\boldsymbol{\delta}_n) = \max_{j=1, \dots, \max\{\delta_1, \dots, \delta_n\}} \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t}))|_{\delta_t=j}$ , and the proportional constant is given by  $C(\boldsymbol{\delta}_n) = \sum_{j=1}^{\max\{\delta_1, \dots, \delta_n\}} W_j \times \prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t}))|_{\delta_t=j} + (1 - \sum_{j=1}^{\max\{\delta_1, \dots, \delta_n\}} W_j) \times M(\boldsymbol{\delta}_n)$ . Then simulate a Uniform(0, 1) random variable  $U_i$  and accept the proposal to move to  $\boldsymbol{\delta}_n(i, j) = \{\delta_1, \dots, \delta_{i-1}, \delta_i = j, \delta_{i+1}, \dots, \delta_n\}$  if  $j$  satisfies  $\sum_{\ell=0}^{j-1} q_n(i, \ell) < U_i \leq \sum_{\ell=1}^j q_n(i, \ell)$ , where  $q_n(i, 0) = 0$ . The acceptance probability for this Metropolis-Hastings is given by

$$\alpha(\boldsymbol{\delta}_n, \boldsymbol{\delta}_n(i, j)) = \begin{cases} 1 & \text{if } j \leq \max\{\delta_1, \dots, \delta_n\} \text{ and } \max\{\boldsymbol{\delta}_n(i, j)\} = \max\{\delta_1, \dots, \delta_n\} \\ \min \left\{ 1, \frac{C(\boldsymbol{\delta}_n)}{C(\boldsymbol{\delta}_n(i, j))} \frac{M(\boldsymbol{\delta}_n(i, j))}{\prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t}))} \right\} & \text{if } j \leq \max\{\delta_1, \dots, \delta_n\} \text{ and } \max\{\boldsymbol{\delta}_n(i, j)\} < \max\{\delta_1, \dots, \delta_n\} \\ \min \left\{ 1, \frac{C(\boldsymbol{\delta}_n)}{C(\boldsymbol{\delta}_n(i, j))} \frac{\prod_{t=1}^n \phi(Y_t | 0, \sigma_t^2(\mathbf{Z}_{\delta_t}))|_{\delta_t=j}}{M(\boldsymbol{\delta}_n)} \right\} & \text{if } j > \max\{\delta_1, \dots, \delta_n\}. \end{cases} \quad (5.9)$$

This completes the task for sampling  $\boldsymbol{\delta}_n$ . Finally, similar to Walker [20], sampling  $\mathbf{Z}$  only needs (5.5). That is, for all  $\{j : \delta_t = j\}$ , sample  $Z_j$  from (5.5), otherwise sample  $Z_j$  from  $H(dZ)$  if there is no  $j$  equal to  $\delta_t$ .

## 6. Application to the Standard & Poor's 500 Financial Index

The methodology is illustrated on the daily logarithm returns of the S&P500 (Standard & Poor's 500) financial index, dated from 2006 Jan 03 to 2009 Dec 31. The data contains a total of 1007 trading days and is available from the Yahoo Finance (URL: <http://finance.yahoo.com/>). The log return is defined as  $Y_t = 100(\ln I_t - \ln I_{t-1})$  where  $I_t$  is the index at time  $t$ . The algorithm described in Section 4 is used to estimate the nonparametric mixture models.

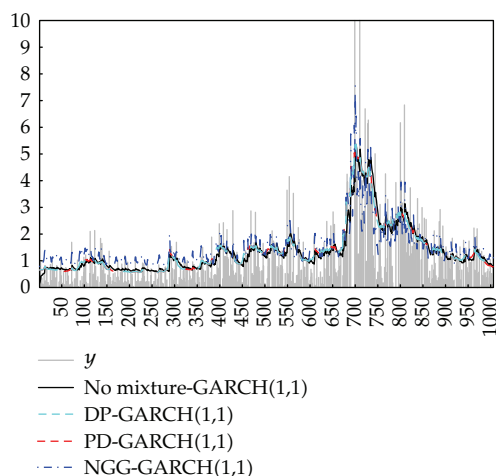
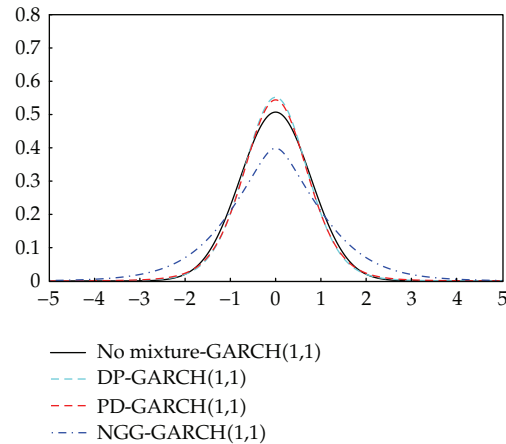


Figure 1: Volatility estimates.

To compare the three different mixture models in Section 3,  $G$  is allocated in the Dirichlet process, the PD process, and the NGG process. In each case, the mean process is denoted  $H$  and is a Gamma-Dirichlet distribution given by  $H(d\nu, d\chi, d\psi) = H_1(d\nu)H_2(d\chi, d\psi)$  where  $H_1(d\nu)$  is the Gamma(1,1) distribution and  $H_2(d\chi, d\psi)$  is the Dirichlet(1,1,1) distribution. We set the parameters of each process such that the variance of each process evaluated over the same measure is equal. This results in  $\theta = 2.3538$  for the Dirichlet process,  $\theta = 0.6769$  and  $\alpha = 1/2$  for the PD and  $\beta = 1$  and  $\alpha = 1/2$  for the NGG. We also compare the results to a no mixture GARCH(1,1) model in which the parameters  $(\nu, \chi, \psi)$  have a prior distribution  $H_1(d\nu)H_2(d\chi, d\psi)$ . We initialise the MCMC algorithm with a partition that separates all integers, that is,  $\mathbf{p}_n = \{C_1 = \{1\}, C_2 = \{2\}, \dots, C_n = \{n\}\}$ . We run the MCMC algorithm for 20,000 iterations of which the first 10,000 iterations are discarded. The last 10,000 iterations are considered a sample from the posterior density of  $\{\mathbf{Z}_n^*, \mathbf{p}_n\}$ .

Figure 1 contains the volatility estimates (fitted data) for the no-mixture model, the Dirichlet process, the DP process, and the NGG process. The no mixture model, the Dirichlet process, and the PD process appear to give similar results. However, it is easy to distinguish the NGG process from the other models since the volatility estimates of the NGG process appear to better fit the observed spikes in the data. Figure 2 presents the predictive densities for each model. Again, the no-mixture model, the Dirichlet process, and the PD process give similar predictive density estimates in the sense that the distribution tails are all similar. However, the NGG process model estimates a predictive density with substantially wider tails than the other three models. Figures 1 and 2 suggest that the Dirichlet and PD processes allocate fewer clusters and consider the periods of increased volatility as outliers within the data. On the other hand the NGG process allocates more clusters and incorporates the periods of increased volatility directly into its predictive density.

Finally, we evaluate the goodness of fit in terms of the marginal likelihoods. The logarithm of the marginal likelihoods of the no mixture model, the Dirichlet process model, PD process model, and the NGG process model are  $-1578.085$ ,  $-1492.086$ ,  $-1446.275$ , and  $-1442.269$ , respectively. Under the marginal likelihood criterion all three mixture models outperform the GARCH(1,1). Further, the NGG process outperforms the PD process which



**Figure 2:** Predictive density estimates.

in turn outperforms the model proposed in Lau and So [9]. These results suggest that generalisations of the Dirichlet process mixture model should be further investigated for time-dependent data.

## 7. Conclusion

In this paper we have extended nonparametric mixture modelling for GARCH models to the Kingman-Poisson process. The process includes the previously applied Dirichlet process and also includes the Poisson-Dirichlet and Normalised Generalised Gamma process. The Poisson-Dirichlet and Normalised Generalised Gamma process provide richer clustering structures than the Dirichlet process and have not been previously adapted to time series data. An application to the S&P500 financial index suggests that these more general random probability measures can outperform the Dirichlet process. Finally, we developed an MCMC algorithm that is easy to implement which we hope will facilitate further investigation into the application of nonparametric mixture models to time series data.

## References

- [1] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [3] L. Bauwens, A. Preminger, and J. V. K. Rombouts, "Theory and inference for a Markov switching GARCH model," *Econometrics Journal*, vol. 13, no. 2, pp. 218–244, 2010.
- [4] Z. He and J. M. Maheu, "Real time detection of structural breaks in GARCH models," *Computational Statistics and Data Analysis*, vol. 54, no. 11, pp. 2628–2640, 2010.
- [5] R. T. Baillie and C. Morana, "Modelling long memory and structural breaks in conditional variances: an adaptive FIGARCH approach," *Journal of Economic Dynamics and Control*, vol. 33, no. 8, pp. 1577–1592, 2009.
- [6] M. Haas, S. Mitnik, M. Paolella, Z. He, and J. Maheu, "Mixed normal conditional heteroskedasticity," *Journal of Financial Econometrics*, vol. 2, pp. 211–250, 2004.

- [7] M. Haas, S. Mittnik, M. Paolella, Z. He, and J. Maheu, "A new approach to markov- switching GARCH models," *Journal of Financial Econometrics*, vol. 2, pp. 493–530, 2004.
- [8] S. Kaufmann and S. Frühwirth-Schnatter, "Bayesian analysis of switching arch models," *Journal of Time Series Analysis*, vol. 23, no. 4, pp. 425–458, 2002.
- [9] J. W. Lau and M. K. P. So, "A Monte Carlo Markov chain algorithm for a class of mixture time series models," *Statistics and Computing*, vol. 21, no. 1, pp. 69–81, 2011.
- [10] M. J. Jensen and J. M. Maheu, "Bayesian semiparametric stochastic volatility modeling," *Journal of Econometrics*, vol. 157, no. 2, pp. 306–316, 2010.
- [11] J. E. Griffin, "Inference in infinite superpositions of Non-Gaussian Ornstein-Uhlenbeck processes using Bayesian nonparametric methods," *Journal of Financial Econometrics*, vol. 9, no. 3, Article ID nbq027, pp. 519–549, 2011.
- [12] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, vol. 1, pp. 209–230, 1973.
- [13] A. Y. Lo, "On a class of Bayesian nonparametric estimates. I. Density estimates," *Annals of Statistics*, vol. 12, no. 1, pp. 351–357, 1984.
- [14] J. Pitman and M. Yor, "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator," *Annals of Probability*, vol. 25, no. 2, pp. 855–900, 1997.
- [15] J. Pitman, "Poisson-Kingman partitions," in *Statistics and Science: A Festschrift for Terry Speed*, vol. 40 of *IMS Lecture Notes Monogr*, pp. 1–34, Beachwood, Ohio, USA, 2003.
- [16] J. F. C. Kingman, "Random discrete distribution," *Journal of the Royal Statistical Society, Series B*, vol. 37, pp. 1–22, 1975.
- [17] L. F. James, "Poisson process partition calculus with applications to exchangeable models and Bayesian nonparametrics," <http://arxiv.org/abs/math/0205093>.
- [18] L. F. James, "Bayesian poisson process partition calculus with an application to Bayesian lévy moving averages," *Annals of Statistics*, vol. 33, no. 4, pp. 1771–1799, 2005.
- [19] A. Lijoi, I. Prünster, and S. G. Walker, "Bayesian nonparametric estimators derived from conditional Gibbs structures," *Annals of Applied Probability*, vol. 18, no. 4, pp. 1519–1547, 2008.
- [20] S. G. Walker, "Sampling the Dirichlet mixture model with slices," *Communications in Statistics*, vol. 36, no. 1, pp. 45–54, 2007.
- [21] O. Papaspiliopoulos and G. O. Roberts, "Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models," *Biometrika*, vol. 95, no. 1, pp. 169–186, 2008.
- [22] A. Lijoi, I. Prünster, and S. G. Walker, "Investigating nonparametric priors with Gibbs structure," *Statistica Sinica*, vol. 18, no. 4, pp. 1653–1668, 2008.
- [23] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, 1997.
- [24] Z. Zhang, W. K. Li, and K. C. Yuen, "On a mixture GARCH time-series model," *Journal of Time Series Analysis*, vol. 27, no. 4, pp. 577–597, 2006.
- [25] W. Vervaat, "On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables," *Advances in Applied Probability*, vol. 11, pp. 750–783, 1979.
- [26] A. Brandt, "The stochastic equation  $Y_{n+1} = A_n Y_n + B_n$  with stationary coefficients," *Advances in Applied Probability*, vol. 18, no. 1, pp. 211–220, 1986.
- [27] J. K. Ghosh and R. V. Ramamoorthi, *Bayesian Nonparametrics*, Springer Series in Statistics, Springer, New York, NY, USA, 2003.
- [28] J. Pitman, *Combinatorial Stochastic Processes*, vol. 1875 of *Lecture Notes in Mathematics*, Springer, Berlin, Germany, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [29] M. West, P. Müller, and M. D. Escobar, "Hierarchical priors and mixture models, with applications in regression and density estimation," in *A Tribute to D. V. Lindley*, pp. 363–386, John Wiley & Sons, New York, NY, USA, 1994.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

