A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model With Application to Portfolio Selection

Audronė Virbickaitė

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

audrone.virbickaite@uc3m.es

M. Concepción Ausín

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

concepcion.ausin@uc3m.es

Pedro Galeano

Universidad Carlos III de Madrid

Getafe (Madrid), Spain, 28903

pedro.galeano@uc3m.es

Abstract

We propose a Bayesian non-parametric approach for modeling the distribution of multiple returns. In particular, we use an asymmetric dynamic conditional correlation (ADCC) model to estimate the time-varying correlations of financial returns where the individual volatilities are driven by GJR-GARCH models. The ADCC-GJR-GARCH model takes into consideration the asymmetries in individual assets' volatilities, as well as in the correlations. The errors are modeled using a Dirichlet location-scale mixture of multivariate Gaussian distributions allowing for a great flexibility in the return distribution in terms of skewness and kurtosis. Model estimation and prediction are developed using MCMC methods based on slice sampling techniques. We carry out a simulation study to illustrate the flexibility of the proposed approach. We find that the proposed DPM model is able to adapt to several frequently used distribution models and also accurately estimates the posterior distribution of the volatilities of the returns, without assuming any underlying distribution. Finally, we present a financial application using Apple and NASDAQ Industrial index data to solve a portfolio allocation problem. We find that imposing a restrictive parametric distribution can result into underestimation of the portfolio variance, whereas DPM model is able to overcome this problem.

Keywords: Bayesian Analysis; Dirichlet Process Mixtures; Markov Chain Monte Carlo; Multivariate GARCH; Portfolio Allocation.

JEL Classification: C11, C32, C53, C58, G11.

1. INTRODUCTION

Modeling the stylized features of the assets' returns has been extensively researched for decades and the topic yet remains of great interest. The ARCH-family models, first introduced by Engle (1982) and then generalized by Bollerslev (1986), are without doubt the most analyzed and used in practice to explain time-varying volatilities, see Bollerslev et al. (1992), Bollerslev et al. (1994), Engle (2002b), Teräsvirta (2009) and Tsay (2010), for instance. When dealing with multiple returns, one must also take into consideration the mutual dependence between them, see Bauwens et al. (2006), Silvennoinen & Teräsvirta (2009) and Tsay (2010), for instance. In particular, conditional correlation models firstly proposed by Engle (2002a), Tse & Tsui (2002) and Christodoulakis & Satchell (2002), play an important role because there is evidence that conditional correlations are time dependent. The asymmetric behavior of individual returns has been well established in the financial literature, see Hentschel (1995). More recently, Cappiello et al. (2006) have pointed out that conditional correlations may exhibit some stylized features, such as persistence and asymmetry and, consequently, have proposed Asymmetric Dynamic Conditional Correlation (ADCC) models for time-varying correlations.

It is well known, that every prediction, in order to be useful, has to come with a certain precision measurement. In this way the agent can know the uncertainty of the risk she is facing. In the field of MGARCH models, the distribution of the returns, that strongly depends on the distributional assumptions for the error term, permits to quantify this uncertainty about the future. However, the traditional premises of multivariate Normal or Student-t distributions may be rather restrictive because usually the empirical distribution of returns is slightly skewed and their tails are fatter than those of a Gaussian distribution, see Rossi & Spazzini (2010), for example. Alternative parametric choices such as the Student-t density, see Fiorentini et al. (2003), the skew-Student-t distribution, see Bauwens & Laurent (2005), or finite mixtures of Gaussian distributions, see e.g. Ausín & Galeano (2007) and

Galeano & Ausín (2010), have been proposed in the literature that usually improve the fit of the GARCH models. However, all of them require the assumption of a certain parametric model.

In this paper, in order to model the stylized features of the assets' returns, we assume a model which allows for asymmetries not only in individual assets' returns, but also in their correlations. In particular, we consider an Asymmetric Dynamic Conditional Correlation (ADCC) model for time-varying correlations, proposed by Cappiello et al. (2006), with GJR-GARCH models, proposed by Glosten et al. (1993), for individual volatilities. This specification, denoted by ADCC-GJR-GARCH, provides a much more realistic evaluation of the co-movements of the assets' returns than standard symmetric MGARCH models. Additionally, we propose a Bayesian non-parametric approach for the ADCC-GJR-GARCH model avoiding the specification of a particular parametric distribution for the return innovations. More specifically, we consider a Dirichlet Process Mixture (DPM) model of Gaussian distributions, firstly introduced by Antoniak (1974). This is a very flexible model that can be viewed as an infinite location-scale mixture of Gaussian distributions which includes, among others, the Gaussian, Student-t, logistic, double exponential, Cauchy and generalized hyperbolic distributions, among others. We follow closely the works of Ausín et al. (2014), who have applied the DPM models for univariate GJR-GARCH, and Jensen & Maheu (2013), who have used DPM models for the multivariate symmetric DVEC by Ding & Engle (2001). Non-parametric time-varying volatility models have been of great interest in the recent literature, not only in GARCH, but also in Stochastic Volatility setting, see Jensen & Maheu (2010), Jensen & Maheu (2012), Delatola & Griffin (2011) and Delatola & Griffin (2013).

The Bayesian approach also helps to deal with parameter uncertainty in portfolio decision problems, see e.g. Jorion (1986), Greyserman et al. (2006), Avramov & Zhou (2010) and Kang (2011), among others. This is in contrast with the usual maximum likelihood estimation approach, which assumes a "certainty equivalence" viewpoint, where the sample estimates are treated as the true values, which is not always correct and has been criticized in a number

of papers. As noted by Jorion (1986), this estimation error can gravely distort optimal portfolio selection. In this paper, we propose a Bayesian method which provides the posterior distributions of the one-step-ahead optimal portfolio weights, which are more informative than simple point estimates. In particular, using the proposed approach, it is possible to obtain Bayesian credible intervals for the optimal portfolio weights. Note that the Bayesian methodology also provides some other advantages over the classical maximum likelihood techniques, see Ardia & Hoogerheide (2010). For example, it is easy to incorporate via priors complicated positivity constraints on the parameters to ensure positive variance and covariance stationarity. Additionally, it is possible to approximate the posterior distribution of any other non-linear function of the parameters, as will be done for the optimal portfolio weights. For a survey on Bayesian inference methods for univariate and multivariate GARCH models see Virbickaite et al. (2013).

Therefore, the main contribution of this work is the proposal of a Bayesian nonparametric method for explaining the dynamics of the assets' returns via an ADCC-GJR-GARCH model with the use of DPM location-scale mixture models for the return innovations. Also, we present an application of Bayesian non-parametric techniques in portfolio decision problems and explore the differences in uncertainty between the proposed approach and conventional restrictive distributional assumptions, where the objective is to provide a more realistic evaluation of risk of financial decisions. As commented before, this study extends the work by Ausín et al. (2014) to the multivariate framework and the recent work by Jensen & Maheu (2013) to the asymmetric setting. Also, differently from the work of Jensen & Maheu (2013), we always assume a conjugate prior specification and we use a different sampling approach.

The outline of the paper is as follows: Section 2 describes the model, inference and prediction from a Bayesian perspective. Section 3 introduces the time-varying portfolio optimization problem. Section 4 presents a short simulation study. Section 5 illustrates the proposed approach using a real data example, solves a portfolio allocation problem and carries out model comparison. Finally, Section 6 concludes.

2.   MODEL, INFERENCE AND PREDICTION

This section describes the asymmetric dynamic conditional correlation GJR-GARCH model used for volatilities and correlations. Then, we introduce the DPM specification for the error term. Finally, we provide a detailed explanation of the implementation of Bayesian non-parametric inference and the methodology of obtaining predictive densities of returns and volatilities.

2.1   The ADCC-GJR-GARCH-DPM Model

As commented before, financial returns usually exhibit two types of asymmetries: in individual volatilities and in conditional correlations. Therefore, in one hand, we choose the GJR-GARCH model proposed by Glosten et al. (1993) for individual returns, to incorporate asymmetric volatility effects, while, on the other hand, we use the ADCC model proposed by Cappiello et al. (2006) and based on the previous work by Engle (2002a), to model joint volatilities. Then, we assume that the vector of K asset returns is given by:

$$r_t = H_t^{1/2} \epsilon_t, \tag{1}$$

for $t = 1, \ldots, T$, where H_t is a scale symmetric $K \times K$ matrix and ϵ_t are a sequence of iid random variables with an unknown K-dimensional distribution $\mathcal{F}_K$, for which we will assume a DPM prior specified later. As usual in all the DCC models, the matrix H_t can be decomposed as follows:

$$H_t = D_t R_t D_t, \tag{2}$$

where D_t is a diagonal matrix that contains the square root of the elements of the main diagonal of H_t, denoted by d_{iit}, for $i = 1, \ldots, K$. These elements are assumed to follow

GJR-GARCH models given by:

$$d_{iit}^2 = \omega_i + (\alpha_i + \phi_i I_{i,t-1}) r_{it-1}^2 + \beta_i d_{iit-1}^2, \tag{3}$$

with parameters $\omega_i, \alpha_i, \phi_i, \beta_i > 0$, for $i = 1, \ldots, K$ and where $I_{i,t-1}$ is an indicator function such that $I_{i,t-1} = 1$ if $r_{it} < 0$ and is $I_{i,t-1} = 0$, otherwise. On the other hand, to introduce R_t, we need to define

$$\varepsilon_t = D_t^{-1} r_t, \text{ and } \eta_t = \varepsilon_t \odot I(\varepsilon_t < 0), \tag{4}$$

where $\odot$ denotes Hadamard matrix product operator and $I(\varepsilon_t < 0)$ is a vector with ith component equal to 1 if $\varepsilon_{it} < 0$, and 0 otherwise. Then, R_t is given by:

$$R_t = Q_t^{\star-1} Q_t Q_t^{\star-1}, \tag{5}$$

where Q_t is the $K \times K$ matrix given by:

$$Q_t = S(1 - \kappa - \lambda - \delta/2) + \kappa \times \varepsilon_{t-1} \varepsilon_{t-1}' + \lambda \times Q_{t-1} + \delta \times \eta_{t-1} \eta_{t-1}', \tag{6}$$

where S is a sample correlation matrix of ε_t, and $Q_t^\star$ is a diagonal matrix with the square root of the ith diagonal element of Q_t on its ith diagonal position. We impose that $\kappa, \lambda, \delta > 0$ and $\kappa + \lambda + \delta/2 < 1$ to ensure the positivity and stationarity of Q_t. Finally, the vector $\Phi = (\omega, \alpha, \beta, \phi, \kappa, \lambda, \delta)$ summarizes the set of parameters describing the matrices H_t, for $t = 1, 2, \ldots$

As for the unknown distribution of $\epsilon_t \sim \mathcal{F}_K$, there has been and ongoing discussing about the best specification for the heavy-tailed financial returns. Next, we present a flexible DPM specification for the errors and some of the most important special cases arising from this model. Using the stick-breaking representation by Sethuraman (1994), a DPM of Gaussian distributions can be expressed as a location-scale Gaussian mixture model with infinitely many components and therefore, it can be easily defined as an extension of a parametric

mixture model. The base distribution of the DP, usually denoted by G_0, corresponds to the prior distribution of the component parameters in the infinite mixture. The concentration parameter, denoted by c, can be interpreted as the prior belief on the number of clusters in the mixture. Small values of c assumes a priori an infinite mixture model with a small number of components with large weights. On the contrary, large values of c assumes a priori an infinite mixture model with all the weights being very small.

Therefore, the resulting density function of ϵ_t can be written as:

$$f\left(\epsilon_t | \rho, \mu, \Lambda\right) = \sum_{j=1}^{\infty} \rho_j \mathcal{N}_K \left(\epsilon_t | \mu_j, \Lambda_j^{-1}\right), \tag{7}$$

where $\mathcal{N}_K$ denotes a K-variate normal density. Let us denote by $\Omega = \{\rho_j, \mu_j, \Lambda_j\}_{j=1}^{\infty}$ the infinite-dimensional parameter vector describing the innovation mixture distribution. Here ρ_j, represent the component weights, μ_j are the component means and Λ_j are the precision matrices, for $j = 1, 2, \ldots$ Using the stick breaking representation, the weights of the infinite mixture components are reparameterized as follows: $\rho_1 = v_1$, $\rho_j = (1 - v_1) \ldots (1 - v_{j-1}) v_j$, where a Beta prior distribution is assumed for $v_j \sim \mathcal{B}(1, c)$, for $j = 1, 2, \ldots$ Clearly, there will be some sensitivity to the choice of the concentration parameter c. Therefore, we further assume a Gamma hyper-prior distribution for c, $c \sim \mathcal{G}(a_0, b_0)$. Finally, as a base distribution, we assume a conjugate Normal-Wishart prior for $(\mu_j, \Lambda_j) \sim \mathcal{NW}(m_0, s_0, W_0, d_0)$, where:

$$\mu_j | \Lambda_j \sim \mathcal{N}_K \left(m_0, (s_0 \Lambda_j)^{-1}\right),$$

$$\Lambda_j \sim \mathcal{W}(W_0, d_0),$$

for $j = 1, 2, \ldots$, such that $\mathrm{E}\left[\Lambda_j\right] = d_0 \times W_0^{-1}$ and $\mathrm{E}\left[\Lambda_j^{-1}\right] = (d_0 - (K + 1)/2)^{-1} \times W_0$.

In summary, the complete set of model parameters is denoted by $\Theta = (\Phi, \Omega)$. Given the information available up to time $t - 1$, denoted by r^{t-1}, the conditional density of the returns

can be written as follows:

$$f(r_t|\Theta, r^{t-1}) = \sum_{j=1}^{\infty} \rho_j \mathcal{N}_K \left(r_t | H_t^{1/2} \mu_j, H_t^{1/2} \Lambda_j^{-1} (H_t^{1/2})' \right), \tag{8}$$

with conditional mean given by:

$$\mu_t^{\star} = E \left[r_t | \Theta, r^{t-1} \right] = H_t^{1/2} \sum_{j=1}^{\infty} \rho_j \mu_j \tag{9}$$

and conditional covariance matrix:

$$H_t^{\star} = \text{Cov} \left[r_t | \Theta, r^{t-1} \right] = H_t^{1/2} \text{Cov} \left[\epsilon_t | \Omega \right] (H_t^{1/2})', \tag{10}$$

where

$$\text{Cov} \left[\epsilon_t | \Omega \right] = \sum_{j=1}^{\infty} \rho_j \left(\Lambda_j^{-1} + \mu_j (\mu_j)' \right) - \left(\sum_{j=1}^{\infty} \rho_j \mu_j \right) \left(\sum_{j=1}^{\infty} \rho_j \mu_j \right)'.$$

It is important to notice that this full unrestricted model induces GARCH-in-Mean effects, since the conditional mean of the returns is not restricted to be zero. Moreover, the DPM model for ϵ_t does not assume an identity covariance matrix. As noted in Jensen & Maheu (2013), imposing moment restrictions in DPM models is still an open question. However, the prior information considered essentially centers ϵ_t around an identity covariance matrix.

On the other hand, an essential issue in choosing more complicated models versus the simple ones is the ability to handle numerous assets. The DPM model is very flexible in this sense, since the general specification described before contains numerous other simplified models. For example, it clearly contains the single Gaussian as a special case when the first mixture weight is equal to one. Also, it is possible to impose a symmetric distribution for the innovations by simply assuming that the mixture means are all equal and, in particular, it could be reasonable to impose $\mu_j = 0$, for $j = 1, 2, \ldots$. If we further assume that the precision

7

matrices are all diagonal, $\Lambda_j = \text{diag}(\lambda_{j1}, \ldots, \lambda_{jK})$, this will lead to uncorrelated innovations. Finally, we could in addition assume that the diagonal elements of each precision matrix are all equal by considering $\Lambda_j = \lambda_j I_K$. In this paper we will use the full version of the DPM model to illustrate the flexibility of it. However, the adaptation of the model to these particular cases in order to simplify the problem of many assets is straightforward.

2.2 MCMC algorithm

The following section describes a Markov Chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution of the parameters of the ADCC-GJR-GARCH-DPM model introduced in the previous section. The algorithm is based on the procedure by Walker (2007), who introduces slice sampling schemes to deal with the infiniteness in DPM, the retrospective MCMC method of Papaspiliopoulos & Roberts (2008) and the ideas by Papaspiliopoulos (2008) who combines these two methods to obtain a new composite algorithm, which is better, faster and easier to implement. Generally, all these approaches compared to traditional schemes based on the original algorithm by Escobar & West (1995) produce better mixing and simpler algorithms.

Following Walker (2007), in order not to sample an infinite number of values at each MCMC step, we introduce a latent variable u_t, such that the joint density of (ϵ_t, u_t) given Ω is:

$$f(\epsilon_t, u_t|\Omega) = \sum_{j=1}^{\infty} \mathbb{1}(u_t < \rho_j) \mathcal{N}_K(\epsilon_t|\mu_j, \Lambda_j^{-1}). \tag{11}$$

Let $A_\rho(u_t) = \{j : \rho_j > u_t\}$ be a set of size N_{u_t}, which is finite for all $u_t > 0$. Then the joint density of (ϵ_t, u_t) in (11) can be equivalently written as $f(\epsilon_t, u_t|\Omega) = \sum_{j \in A_\rho(u_t)} \mathcal{N}_K(\epsilon_t|\mu_j, \Lambda_j^{-1})$. Integrating over u_t gives us the density of infinite mixture of distributions (7). Finally, given u_t, the number of mixture components is finite. In order to simplify the likelihood, we also need to introduce a further indicator latent variable z_t, which indicates the mixture component that ϵ_t comes from: $f(\epsilon_t, z_t = j, u_t|\Omega) = \mathcal{N}_K(\epsilon_t|\mu, \Lambda^{-1}) \mathbb{1}(j \in A_\rho(u_t))$. Then, the

log-likelihood of Θ, given the latent variables u_t and z_t looks as follows:

$$\log L(\Theta|\{r_t, u_t, z_t\}_{t=1}^T) = -\frac{1}{2}\sum_{t=1}^T \left(K\log(2\pi) + \log|H^\star_{t,z_t}| + (r_t - \mu^\star_{t,z_t})H^{\star-1}_{t,z_t}(r_t - \mu^\star_{t,z_t})'\right),$$

$$(12)$$

where $\mu^\star_{t,z_t}$ and $H^\star_{t,z_t}$ are the conditional mean vector and conditional covariance matrix given z_t, i.e.:

$$\mu^\star_{t,z_t} = H_t^{1/2}\mu_{z_t},$$

$$H^\star_{t,z_t} = H_t^{1/2}\Lambda_{z_t}^{-1}H_t^{1/2}.$$

respectively.

Using these latent variables, we now construct the following MCMC algorithm that is described step by step.

Firstly, given z_t, for $t = 1, \ldots, T$, the conditional posterior distribution of the concentration parameter, c, is independent of the rest of the parameters, as seen in Escobar & West (1995). So, we first sample an auxiliary variable $\xi \sim \mathcal{B}(c+1, T)$ and then c from a Gamma mixture:

$$\pi_\xi \mathcal{G}(a_0 + z^\star, b_0 - \log(\xi)) + (1 - \pi_\xi)\mathcal{G}(a_0 + z^\star - 1, b_0 - \log(\xi)),$$

where $z^\star = \max(z_1, \ldots, z_T)$ and $\pi_\xi = (a_0 + z^\star - 1)/(a_0 + z^\star - 1 + T(b_0 - \log(\xi)))$.

In the second step, we sample from the conditional posterior of v_j for $j = 1, \ldots, z^\star$, which is given by:

$$v_j|\{z_t\}_{t=1}^T \sim \mathcal{B}(n_j + 1, T - \sum_{l=1}^j n_l + c),$$

where n_j is the number of observations in the jth component and $\sum_{l=1}^j n_l$ gives the cumulative sum of the groups. Also, $\rho_1 = v_1$, $\rho_j = (1 - v_1)\ldots(1 - v_{j-1})v_j$, for $j = 2, \ldots, z^\star$.

9

In the third step, we sample the uniform latent variables $u_t \sim \mathcal{U}(0, \rho_{z_t})$, for $t = 1, \ldots, T$.

In the fourth step, following Walker (2007), we need to find the smallest $j^\star$ such that $\sum_{j=1}^{j^\star} \rho_j > u^\star$, where $u^\star = \min(u_1, \ldots, u_T)$. Then, if $z^\star < j^\star$, we need to sample v_j, for $j = z^\star + 1, \ldots, j^\star$, from the prior and sample ρ_j accordingly.

Next, in the fifth step, we sample from the conditional posterior distribution of the mixture parameters, which are also Normal-Wishart distributions, $(\mu_j, \Lambda_j) \sim \mathcal{NW}(m_j, s_j, W_j, d_j)$, for $j = 1, \ldots, j^\star$, where:

$$m_j = \frac{s_0 m_0 + n_j \bar{\epsilon}_j}{s_0 + n_j}, \qquad s_j = s_0 + n_j,$$

$$W_j = W_0^{-1} + S_j + \frac{s_0 n_j}{s_0 + n_j}(m_0 - \bar{\epsilon}_j)(m_0 - \bar{\epsilon}_j)',$$

$$S_j = \frac{1}{n_j} \sum_{t:z_t=j}^{T} (\epsilon_t - \bar{\epsilon}_j)(\epsilon_t - \bar{\epsilon}_j)', \qquad \bar{\epsilon}_j = \frac{1}{n_j} \sum_{t:z_t=j}^{T} \epsilon_t,$$

$$d_j = d_0 + n_j.$$

Note that this approach is different from the one described in Jensen & Maheu (2013) since they assume independent prior distributions for μ_j and Λ_j, and then, it is necessary to include some Gibbs steps to sample from the conditional posterior.

In the sixth step, we generate to which component each observation belongs to by sampling from the following conditional posterior distribution:

$$\Pr(z_t = j | \ldots) \propto \mathbb{1}\left(j \in A_\rho(u_t)\right) \mathcal{N}_K(\epsilon_t | \mu_j, \Lambda_j^{-1}),$$

see also Walker (2007).

The rest of the steps of the algorithm concern updating the parameters of the ADCC-GJR-GARCH model. For that, we use the Random Walk Metropolis Hasting (RWMH), following a similar procedure as in Jensen & Maheu (2013). For the set of parameters Φ a candidate value $\tilde{\Phi}$ is generated from a D-variate Normal distribution with mean equal to the

previous value of the parameter, where $D = 4K + 3$, is the number of parameters in Φ as follows:

$$\tilde{\Phi} \sim \begin{cases} \mathcal{N}_D\left(\Phi, V\right) & \text{w.p.} \quad p \\ \mathcal{N}_D\left(\Phi, 100V\right) & \text{w.p.} \quad 1-p \end{cases}$$

The probability of accepting a proposed value $\tilde{\Phi}$, given the current value Φ, is $\alpha(\Phi, \tilde{\Phi}) = \min\left\{1, \prod_{t=1}^{T} l(r_t|\Phi) / \prod_{t=1}^{T} \tilde{l}(r_t|\tilde{\Phi})\right\}$, where the likelihood used is as in (12), see e.g. Robert & Casella (2005). The covariance matrix V is obtained by running some initial MCMC iterations and then adjusting the sample covariance matrix by some factor in order to achieve the desired acceptance probability. In this paper the acceptance probabilities are adjusted to be between 20% and 50% and we use $p = 0.9$.

2.3   Prediction

In this section, we are mainly interested in estimating the one-step-ahead predictive density of the returns:

$$f(r_{T+1}|r^T) = \int f(r_{T+1}|\Theta, r^T)f(\Theta|r^T)d\Theta, \tag{13}$$

where $f(r_{T+1}|\Theta, r^T)$ is specified in (8). Although this integral is not analytically tractable, it can be in principle approximated using the MCMC output,

$$f(r_{T+1}|r^T) \simeq \frac{1}{M} \sum_{m=1}^{M} f(r_{T+1}|\Theta^{(m)}, r^T), \tag{14}$$

where M is the length of the MCMC chain and $\Theta^{(m)}$ is the infinite set of parameters at the m-th iteration. However, in practice, at each iteration, there are a finite number of weigths, $\rho_j^{(m)}$, for $j = 1, \ldots, j^{\star(m)}$, and the corresponding pairs of means, $\mu_j^{(m)}$, and precision matrices, $\Lambda_j^{(m)}$. Then, as considered in Jensen & Maheu (2013), we can use the following simulation procedure at each MCMC iteration.

Repeat for $r = 1, \ldots, R$:

i. Sample a random variable $a \sim \mathcal{U}(0, 1)$.

ii. Take such $\rho_r^{(m)}$ for which $\sum_{j=1}^{r-1} \rho_{j-1}^{(m)} < a < \sum_{j=1}^{r} \rho_j^{(m)}$ and the corresponding $(\mu_r, \Lambda_r)^{(m)}$.

iii. If $\sum_{j=1}^{j^{\star(m)}} \rho_j^{(m)} < a$, sample $(\mu_r, \Lambda_r)^{(m)}$ from the Normal-Wishart prior.

And then, approximate the one-step-ahead density in (14) by,

$$f(r_{T+1}|\Theta^{(m)}, r^T) \simeq \frac{1}{R} \sum_{r=1}^{R} \mathcal{N}_K \left(r_{T+1}|\mu_r^{(m)} H_{T+1}^{(m)1/2}, H_{T+1}^{(m)1/2} \left(\Lambda_r^{(m)}\right)^{-1} (H_{T+1}^{(m)1/2})' \right), \qquad (15)$$

where $(\mu_r, \Lambda_r)^{(m)}$, are the R pairs of means and precision matrices simulated for $r = 1, \ldots, R$, and $H_{T+1}^{(m)}$ is the value of the H_{T+1} matrix at the m-th MCMC iteration.

Using this simulation procedure, we can also obtain predictions for many other important measures. For example, the posterior expected value of the adjusted one-step-ahead mean and volatility matrix, introduced in (9) and (10), respectively, can be approximated by:

$$\mathrm{E}\left[\mu_{T+1}^{\star} \mid r^T\right] \simeq \frac{1}{M} \sum_{m=1}^{M} \mu_{T+1}^{\star(m)}, \qquad (16)$$

and

$$\mathrm{E}\left[H_{T+1}^{\star} \mid r^T\right] \simeq \frac{1}{M} \sum_{m=1}^{M} H_{T+1}^{\star(m)}, \qquad (17)$$

where,

$$\mu_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2} \left(\frac{1}{R} \sum_{r=1}^{R} \mu_r^{(m)}\right) \qquad (18)$$

and

$$H_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2} \left(\frac{1}{R} \sum_{r=1}^{R} \left((\Lambda_r^{(m)})^{-1} + \mu_r^{(m)}(\mu_r^{(m)})'\right) - \left(\frac{1}{R} \sum_{r=1}^{R} \mu_r^{(m)}\right) \left(\frac{1}{R} \sum_{r=1}^{R} \mu_r^{(m)}\right)'\right) \left(H_{T+1}^{(m)1/2}\right)'. \qquad (19)$$

In order to obtain the posterior distributions of the adjusted means and volatilities, one should fix a certain R: a number of components to sample at each iteration m. Since the number of components in the data is not known a priori, one might chose R depending on the number of clusters in the data. However, this implies that there is no upper limit for R, which might result into sampling a very large number of components at each step and increasing computational cost. Instead, we propose to make use of the truncation introduced in (11) and use only $j^\star$ components with their corresponding weights ρ at each step, such that equations (18) and (19) become the following:

$$\mu_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2} \left(\sum_{j=1}^{j^{\star(m)}} \rho_i^{(m)} \mu_i^{(m)} \right) \tag{20}$$

and

$$H_{T+1}^{\star(m)} = H_{T+1}^{(m)1/2} \times \tag{21}$$
$$\left(\sum_{j=1}^{j^{\star(m)}} \rho_i^{(m)} \left(\left(\Lambda_i^{(m)} \right)^{-1} + \mu_i^{(m)} (\mu_i^{(m)})' \right) - \left(\sum_{j=1}^{j^{\star(m)}} \rho_i^{(m)} \mu_i^{(m)} \right) \left(\sum_{j=1}^{j^{\star(m)}} \rho_i^{(m)} \mu_i^{(m)} \right)' \right) \left(H_{T+1}^{(m)1/2} \right)'.$$

Similarly, we can approximate the posterior median and credible intervals using the quantiles of the posterior samples $\left\{ \mu_{T+1}^{\star(m)} \right\}_{m=1}^{M}$ and $\left\{ H_{T+1}^{\star(m)} \right\}_{m=1}^{M}$.

3.   PORTFOLIO DECISIONS

As commented in the introduction, optimal asset allocation is greatly affected by the parameter uncertainty, which has been recognized in a number of papers, see Jorion (1986) and Greyserman et al. (2006), among others. They conclude that in the frequentist setting the estimated parameter values are considered to be the true ones, therefore, the optimal portfolio weights tend to inherit this estimation error. Instead of solving the optimization problem on the basis of the choice of unique parameter values, the investor can choose the

Bayesian approach, because it accounts for parameter uncertainty, as seen in Kang (2011) and Jacquier & Polson (2013), for example.

The main objective of diversification is to reduce investor's exposure to risk. See Markowitz (1952) and Merton (1972) for some classical portfolio optimization references. Nowadays, there is a wide variety of portfolio optimization objectives, such as maximizing agent's utility or minimizing expected shortfall, among many others. In this paper we consider one of the mostly used objectives, where the investor minimizes the portfolio variance. The Global Minimum Variance (GMV) portfolio can be found at the very peak of the efficient frontier. Given the time series of returns $r_1, \ldots, T$, the standard approach is to consider the unconditional covariance matrix of the returns, $\Sigma = \mathrm{Cov}[r_t]$ and solve the following optimization problem:

$$p^\star = \arg \min_{p:p'1_K=1} \mathrm{Var}[r_t^P],$$

where p is the weight vector, 1_K is a K-vector of ones and $r_{T+1}^P = p'r_{T+1}$ is the portfolio return at time point $T + 1$. The problem has solution:

$$p^\star = \frac{\Sigma^{-1}1_K}{1_K'\Sigma^{-1}1_K},$$

that is independent of the time point T. Note that, if we choose to impose the short sale constraint, i.e., $p_i \geq 0, \forall i = 1, \ldots, K$, the problem cannot be solved analytically anymore and it requires numerical optimization techniques.

However, recent results suggest that the use of the time-varying covariance matrix to determine portfolio weights leads to better performing portfolios than the use of a constant covariance matrix. For instance, Giamouridis & Vrontos (2007) find that portfolios, constructed under a dynamic approach, have lower average risk and higher out-of-sample risk-adjusted realized return, see also Yilmaz (2011). Cecchetti et al. (1988) was the first to suggest the use of MGARCH models in optimal allocation context. Since then, there has been a number of papers investigating the differences in estimation and evaluating their

performance using various approaches, from simple OLS, to bivariate vector autorregression (VAR), to GARCH. In particular, Kroner & Sultan (1993), Rossi & Zucca (2002) and Yang & Allen (2005), among others, have shown that GARCH-type models leads to the overall portfolio risk reduction.

Consequentely, to solve the portfolio allocation problem in our case, instead of Σ, we use the adjusted one-step-ahead conditional covariance matrix for the assets returns $H_{T+1}^\star$, defined in (10), which varies continuously on the basis of available information up to time T, r^T. Therefore, we are able to obtain optimal portfolio weights for time point $T + 1$ as follows:

$$p_{T+1}^\star = \frac{H_{T+1}^{\star-1}1_K}{1_K' H_{T+1}^{\star-1}1_K}. \tag{22}$$

Using the MCMC output, we can obtain samples from the entire posterior distribution of optimal portfolio weights for $T+1$, $f(p_{T+1}^\star|r^T)$. This approach relies on solving the allocation problem at every MCMC iteration and approximate for example the posterior mean of the optimal portfolio weights by:

$$\mathrm{E}[p_{T+1}^\star|r^T] = \int p_{T+1}^\star f(\Theta|r^T)d\Theta \approx \frac{1}{M}\sum_{m=1}^{M} p_{T+1}^{\star(m)},$$

where $\left\{p_{T+1}^{\star(m)}\right\}_{m=1}^{M}$ is a posterior sample of optimal portfolio weights obtained from (22) for each value of one-step-ahead conditional covariance matrix of the returns, $\left\{H_{T+1}^{\star(m)}\right\}_{m=1}^{M}$, in the MCMC sample. In other words, since we have assembled M one-step-ahead volatility matrices, we can solve the portfolio allocation problem M times. As in the previous section, we can similarly approximate the posterior median of $p_{T+1}^\star$ and credible intervals by using the quantiles of the sample of optimal portfolio weights. In this manner, we are able to approximate the posterior distribution of the optimal portfolio variance, $\sigma_{T+1}^{2\star}$, and optimal portfolio gain, $g_{T+1}^\star$, using the samples $\left\{(p_{T+1}^{\star'}H_{T+1}^\star p_{T+1}^\star)^{(m)}\right\}_{m=1}^{M}$ and $\left\{(p_{T+1}^\star \mu_{T+1}^{\star'})^{(m)}\right\}_{m=1}^{M}$,

respectively, where

$$p\left(\sigma_{t+1,P}|r^t\right) \sim \left\{(\sigma_{t+1,P})^{(m)}\right\}_{m=1}^{M} = \left\{(p_{t+1}^{\star\prime} H_{t+1}^{\star} p_{t+1}^{\star})^{(m)}\right\}_{m=1}^{M},$$

$$p\left(r_{t+1,P}|r^t\right) \sim \left\{(r_{t+1,P})^{(m)}\right\}_{m=1}^{M} = \left\{(p_{t+1}^{\star} r_{t+1}^{\prime})^{(m)}\right\}_{m=1}^{M}.$$

4. SIMULATION STUDY

The goal of this simulation study is to show the flexibility and adaptability of the DPM specification for the innovations for the ADCC-GJR-GARCH model introduced Section 2. For this, we consider three bivariate time series of 3000 observations simulated from a ADCC-GJR-GARCH model with the following innovation distributions: (a) Gaussian $\mathcal{N}(0, I_2)$; (b) Student-t $\mathcal{T}(I_2, \nu = 8)$; (c) Mixture of two bivariate Normals $0.9\mathcal{N}(0, \sigma_1^2 = 0.8, \sigma_{12} = 0.0849, \sigma_2^2 = 0.9) + 0.1\mathcal{N}(0, \sigma_1^2 = 2.8, \sigma_{12} = -0.7637, \sigma_2^2 = 1.9)$. Notice that these are most frequently used distributional assumptions for the financial return data.

Note that, in the third case, we have chosen larger variances for the second mixture component to allow for the presence of extreme returns but preserving an identity covariance matrix. Then, we estimate all three data sets using the proposed ADCC-GJR-GARCH-DPM model assuming uninformative uniform priors restricted to the stationary region for Φ and setting $m_0 = 0_2$, $s_0 = 0.1$, $d_0 = 5$, $W_0 = I_2/5$, $a_0 = 4$ and $b_0 = 4$. The MCMC algorithm is run for 10,000 burn-in plus 40,000 iterations. The point estimates are not reported in the paper to save space. All parameters were estimated well, with true parameters always inside the 95% credible intervals.

Figure 1 presents the contour plots that compare the true one-step-ahead predictive densities of returns, given the model parameters, with the estimated ones, obtained from (14) by setting $R = 3$. As we can see, the estimated contours of the one-step ahead return densities are very close to the true ones. Note that these contours can be seen as a summary of the estimation results for all 11 parameters of the model $\Phi = (\omega, \alpha, \beta, \phi, \kappa, \lambda, \delta)$ and the

distribution for the error term. Therefore, it seems that the proposed infinite mixture model is a very flexible tool that is able to adjust to rather different return specifications. This is of primary interest because in practice one never knows which is the true error distribution. Therefore, the proposed approach appears to be able to fit adequately several frequently used distributions.

Figure 1. Contour plots of the true and estimated one-step-ahead predictive densities, $f(r_{T+1} \mid r^T)$, for the three simulated data sets.
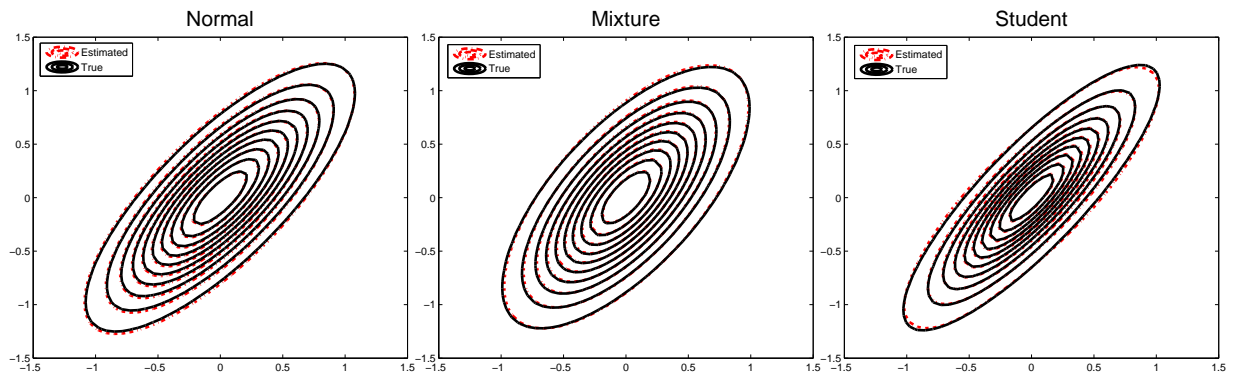
First part of Table 1 presents the estimated posterior mean, median and 95% credible intervals for the number of clusters, $z^\star$, for the three generated datasets. For the Gaussian dataset, the proposed DPM model estimates very few non-empty components, 1.23 on average, where there is always a clear dominant weight. For the Student-t dataset, the proposed DPM model estimates a large number of clusters, around 19.66, with similar small weights. This is expected since, as commented in Jensen & Maheu (2013), the Student-t distribution can be viewed as a limiting case of a DPM model when the concentration parameter goes to infinity and, consequently, the number of clusters increases indefinitely. Finally, for the two-component mixture data, the DPM model can identify very well the two underlying clusters with posterior mean around 2.68. The second and third part of Table 1 also show the estimation results for the concentration parameter, c, and its transformed value $A = c/(1+c)$, where $0 < A < 1$, that has been used by Jensen & Maheu (2013) to provide

Table 1. Posterior mean, median and 95% credible intervals for the number of non-empty clusters, $z^\star$, concentration parameter, c, and quantity A, for the three simulated data sets.

	Gaussian Mean Median	95% CI	Student-t Mean Median	95% CI	2 comp. mixture Mean Median	95% CI
$z^\star$	1.2330 1.0000	(1.0000, 3.0000)	19.6612 19.0000	(9.0000, 33.0000)	2.6765 2.0000	(2.0000, 5.0000)
c	0.3578 0.3252	(0.0934, 0.8072)	2.5037 2.3308	(0.9327, 5.0059)	0.4863 0.4462	(0.1512, 1.0408)
A	0.2509 0.2454	(0.0855, 0.4467)	0.6892 0.6998	(0.4826, 0.8335)	0.3122 0.3085	(0.1314, 0.5100)

an intuition of the probability of having infinite clusters in the mixture. However, note that, different to Jensen & Maheu (2013), we have previously defined a Gamma prior on c instead of a uniform prior on A. Observe that the obtained results are coherent, since the posterior means of c and A for the Gaussian case are the smaller ones ($A = 0.2509$), respectively, while for the Student-t case are the largest ones ($A = 0.6892$), respectively. Finally, for the two-component mixture dataset, the posterior means of c and A are between the corresponding values of the Gaussian and Student-t cases, that can be seen as a compromise between the two extreme cases.

Finally, we have estimated the generated Normal and Student data sets assuming Gaussian and Student's t-distributions, respectively. We used the RWMH with 10,000 burn-in plus 40,000 iterations. This way we were able to obtain a sample of one-step-ahead covariance matrices $\{H_{t+1}^{(m)}\}_{m=1}^M$, estimated using the true return distributions. Figure 2 compares the densities for one-step-ahead covariances $\{H_{t+1}^{\star(m)}\}_{m=1}^M$ assuming a DPM for (a) and (b) with the true data generating model: Gaussian and Student-t respectively. As we can see, the mean estimates and the width and shape of the posterior distributions are very similar for DPM and the ones obtained using the true return distribution. Therefore, we can conclude that DPM model can adjust to different frequently used distributions for the return data without making any restrictive distributional assumptions.

Figure 2. Densities of the elements of the one-step-ahead covariance matrices for Normal and Student data estimated using (a) DPM and Normal and (b) DPM and Student errors.
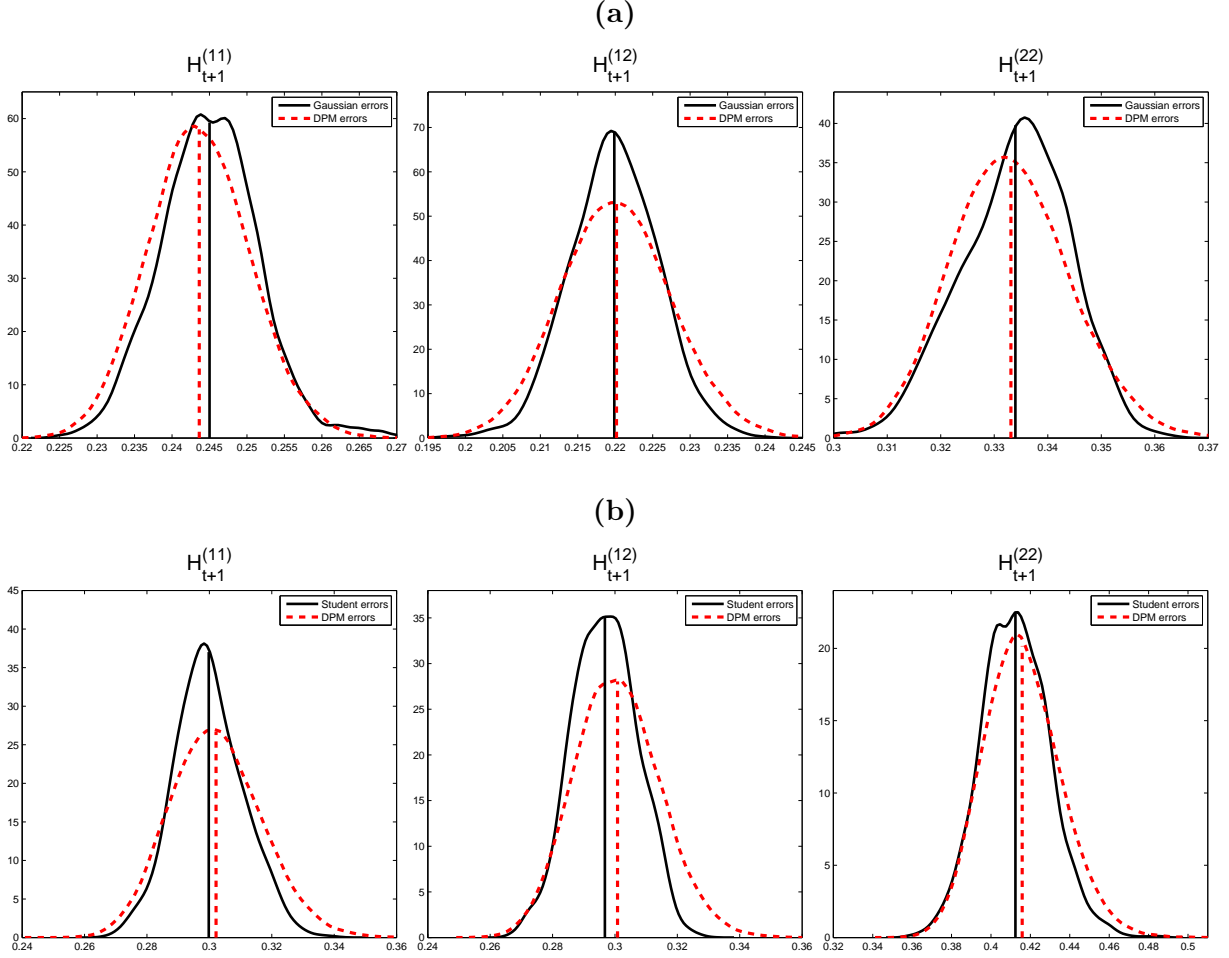
5.   REAL DATA AND RESULTS

In this section, we illustrate the performance of the proposed methodology using a real dataset and solve a portfolio allocation problem as described in Section 3.

5.1   Estimation

We consider the daily price data of Apple Inc. company (P_t^A) and NASDAQ Industrial index (P_t^N) from January 1, 2000 till May 7, 2012, obtained from Yahoo Finance. Then,

Table 2. Descriptive statistics of the Apple Inc. and NASDAQ Ind. return series.

	$100 \times \ln\left(\frac{P_t^A}{P_{t-1}^A}\right)$	$100 \times \ln\left(\frac{P_t^N}{P_{t-1}^N}\right)$
Mean	0.0973	0.0020
Median	0.1007	0.0766
Variance	9.7482	3.1537
Skewness	-4.2492	-0.1487
Kurtosis	102.0411	7.1513
Correlation	0.5376	

daily prices are transformed into daily logarithmic returns (in %), resulting in $T = 3098$ observations. Table 2 provides the basic descriptive statistics and Figure 3 illustrates the dynamics of returns.

Figure 3. Log-returns and histograms of Apple Inc. and NASDAQ Ind. Index.
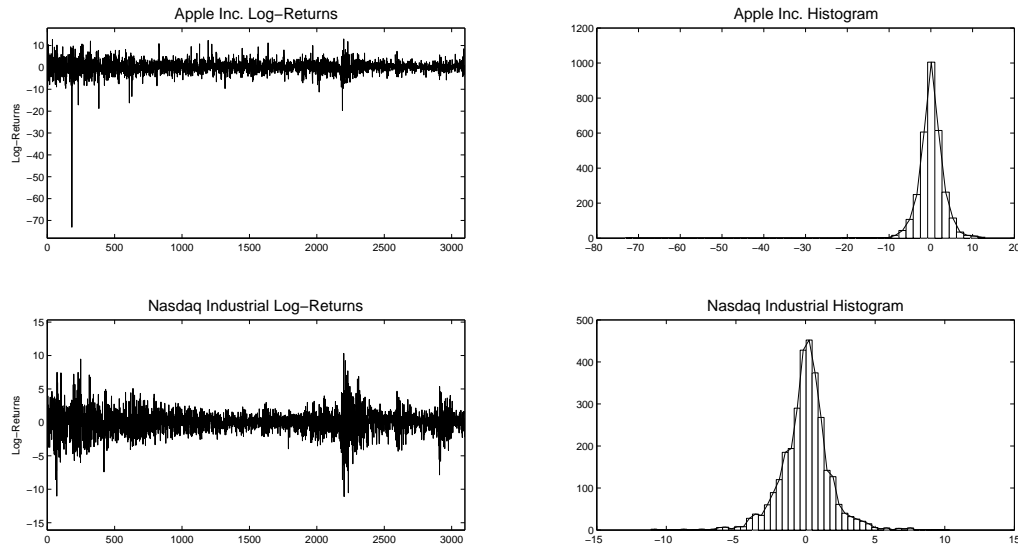
As expected, the Apple Inc. has higher overall variance because of the higher mean return. Both returns do not exhibit any evidence of autoregressive behavior. Apple Inc. returns contain one atypical data point, corresponding to September 29, 2000. The very low return is due to an announcement the day before about lower than expected sales.

Next, the return series was estimated assuming Gaussian, Student-t and DPM errors. Table 3 reports the parameter estimation results for the ADCC model assuming a Gaussian,

a Student-t and the proposed DPM model for the innovation distribution. As we can see from the table, the constant volatility parameter, ω_1, for the first series is overestimated under the Gaussian assumption. This is because the Gaussian model does not allow for fat tails, and therefore, all the volatility is summed into the ω_1. Same happens with the asymmetric volatility parameters ϕ_1 and ϕ_2 under the Gaussian assumption. The posterior mean of the degrees of freedom under the Student-t is around 7. Under the DPM model, the average number of non-empty clusters is $z^\star = 7.8$. Finally, the posterior mean of A is rather close to 0.5, which suggest the better adequacy of the DPM model when compared with the Student-t specification. This is more clear in Figure 4, which draws the histogram of the posterior distribution for A. Note that the posterior probability that A is larger than 0.7 is very small. These results can be compared to the ones obtained in simulation study: when the data comes from a Student-t distribution, DPM estimates a large number of clusters (around 19 in our case) and parameter A is closer to 0.7. On the other hand, when the underlying distribution is Gaussian, DPM model estimates few clusters and parameter A close to 0.2. The results indicate that for this specific data set neither Gaussian, nor Student-t distributions are appropriate, since the data comes from a distribution, which is positioned in between.

Table 3. Estimation results for Apple Inc. (1) and NASDAQ Ind. (2) returns asumming a Gaussian, a Student-t and the proposed DPM model for the innovation distribution.

	Gaussian		Student-t		DPM	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
ω_1	0.2653	(0.1603, 0.3942)	0.1071	(0.0619, 0.1659)	0.1999	(0.1330, 0.2364)
ω_2	0.0285	(0.0203, 0.0395)	0.0192	(0.0125, 0.0269)	0.0088	(0.0052, 0.0131)
α_1	0.0894	(0.0651, 0.1244)	0.0403	(0.0279, 0.0534)	0.0747	(0.0575, 0.0964)
α_2	0.0126	(0.0020, 0.0270)	0.0109	(0.0012, 0.0245)	0.0055	(0.0006, 0.0123)
β_1	0.8430	(0.8039, 0.8740)	0.8975	(0.8730, 0.9204)	0.8771	(0.8489, 0.9000)
β_2	0.9237	(0.9052, 0.9375)	0.9281	(0.9121, 0.9408)	0.9232	(0.9074, 0.9362)
ϕ_1	0.1197	(0.0622, 0.1627)	0.0409	(0.0183, 0.0683)	0.0539	(0.0139, 0.0984)
ϕ_2	0.1050	(0.0796, 0.1312)	0.0739	(0.0520, 0.0912)	0.0370	(0.0255, 0.0476)
κ	0.0093	(0.0030, 0.0266)	0.0075	(0.0020, 0.0153)	0.0172	(0.0043, 0.0326)
λ	0.9828	(0.9415, 0.9936)	0.9711	(0.9477, 0.9858)	0.8914	(0.8445, 0.9409)
δ	0.0061	(0.0008, 0.0213)	0.0220	(0.0094, 0.0392)	0.0242	(0.0024, 0.0426)
ν			7.1879	(7.0529, 7.2862)		
$z^\star$					7.7481	(4.0000, 13.0000)
A					0.4862	(0.2723, 0.6746)

Figure 4. Histogram of the posterior sample of $A = c/(1 + c)$.
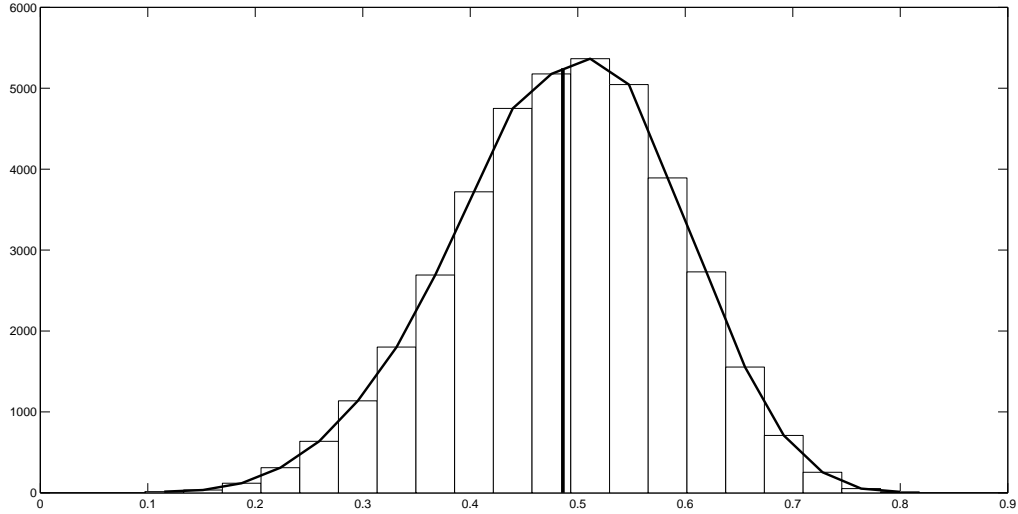

Figure 5 compares the predictive densities of the one-step-ahead return, r_{T+1}, under the Gaussian, the Student-t and the proposed DPM specifications for the innovation distribution. Observe that both Normal and Student-t models lead to symmetric predictive densities, although for the Student-t case it exhibits fatter tails, which are completely defined by a single

parameter, ν. On the contrary, the DPM model, which allows for different variances and non zero means, predicts an asymmetric multimodal density with fatter tails. Also, Figure 6 presents the marginal log predictive densities for the returns under the three specifications. Observe that the DPM model can differentiate between volatile and not so volatile returns, since it predicts obviously fatter tails for the Apple return data, meanwhile for not so volatile NASDAQ data, the difference between DPM and Student-t is not so big. The Gaussian specification in both cases cannot capture the high kurtosis.

Figure 5. Contours of the predictive densities for r_{T+1} under a Gaussian, Student-t and DPM specification for the innovation distribution.
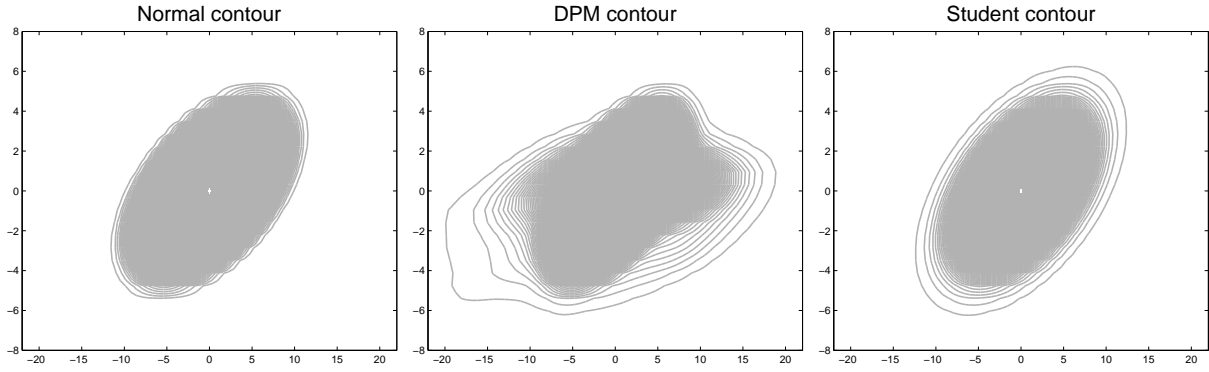

Figure 6. Marginal log-predictive densities for the one-step-ahead Apple (left) and NAS-DAQ (right) return data, under a Gaussian, Student-t and DPM specification for the innovation distribution.
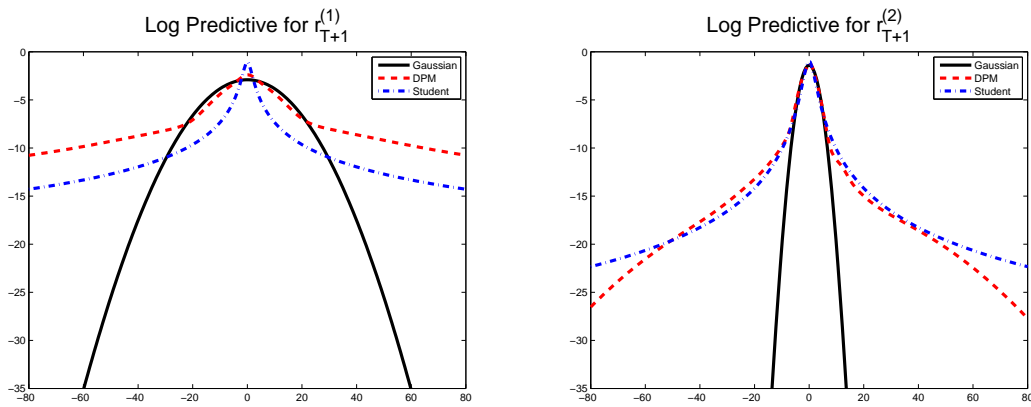
Table 4. Cumulative log-predictive likelihoods for the DPM, Student and Gaussian error models.

Model	$\log p(r_{T+1}, \ldots, r_{T+k})$
DPM	-746.5467
Student	-749.2850
Normal	-775.8299
# of out-of-sample obs.	$k = 233$

Next, following Jensen & Maheu (2013), we compare the three estimated models using predictive likelihoods based on a small set of out-of-sample observations, $\{T + 1, \ldots, T + k\}$. For each new observation, we calculate the predictive likelihood as an average over the MCMC iterations given by:

$$p(r_{T+i}|r^{T+i-1}) = \frac{1}{M} \sum_{m=1}^{M} p(r_{T+i}|r_{T+i-1}, \Theta^{(m)}), \text{ for } i = 1, \ldots, k,$$

and then calculate the sum of the logarithms over the out-of-sample time period:

$$\log p(r_{T+1}, \ldots, r_{T+k}) = \sum_{i=1}^{k} \log p(r_{T+i}|r^{T+i-1}).$$

However, note that, differently to Jensen & Maheu (2013), we do not re-estimate the model whenever a new observation arrives to avoid a high increase in the computational cost, we use the already estimated model parameters up to time T. Table 4 presents the cumulative log-predictive likelihood for the three models using $k = 233$ out-of-sample observations. Observe that the sum of the log-predictive favors the DPM model. The difference between the two competing models is the log of a Bayes factor.

Figure 7 draws the posterior densities of the volatilities and Table 5 presents the posterior means, medians and confidence intervals for the elements of the one-step-ahead volatility matrix under the Gaussian, the Student-t and the proposed DPM specification. For the latter, these are obtained using (17) and the explanations given in Section 2.3. Observe that the credible intervals for the DPM model are wider, especially for the marginal one-step-ahead volatility of the first series. This is because it allows for some very volatile

24

Table 5. Posterior means, medians and confidence intervals for the elements of the one-step-ahead volatility matrix under the Gaussian, Student-t and DPM specification.

	Gaussian		Student-t		DPM	
	Mean Median	95% CI	Mean Median	95% CI	Mean Median	95% CI
$H_{T+1}^{\star(1,1)}$	7.3315 7.3419	(6.8461, 7.6932)	6.6228 6.6237	(6.2307, 7.0214)	6.3519 6.3532	(5.4437, 7.3185)
$H_{T+1}^{\star(1,2)}$	1.7954 1.7890	(1.6379, 1.9831)	1.6684 1.6649	(1.5609, 1.7944)	1.6215 1.6284	(1.3615, 1.8238)
$H_{T+1}^{\star(2,2)}$	1.5932 1.5954	(1.4534, 1.7163)	1.6524 1.6541	(1.5188, 1.7733)	1.4570 1.4634	(1.2624, 1.5953)
ν			7.1879 7.1933	(7.0529, 7.2865)		
$\mu_{T+1}^{\star(1)}$	0		0		0.1503 0.1503	(0.0547, 0.2456)
$\mu_{T+1}^{\star(2)}$	0		0		0.0131 0.0132	(-0.0311, 0.0574)

mixture components, due to the atypical data point, which seems to provide a more realistic evaluation of risk for an agent. Also, under the DPM model, the posterior distributions for the volatilities are not symmetric. Finally, Table 5 also shows the estimation results for the degrees of freedom parameter under the Student-t assumption, which indicates heavy tails, and for the one-step-ahead mean under the DPM model using (16).

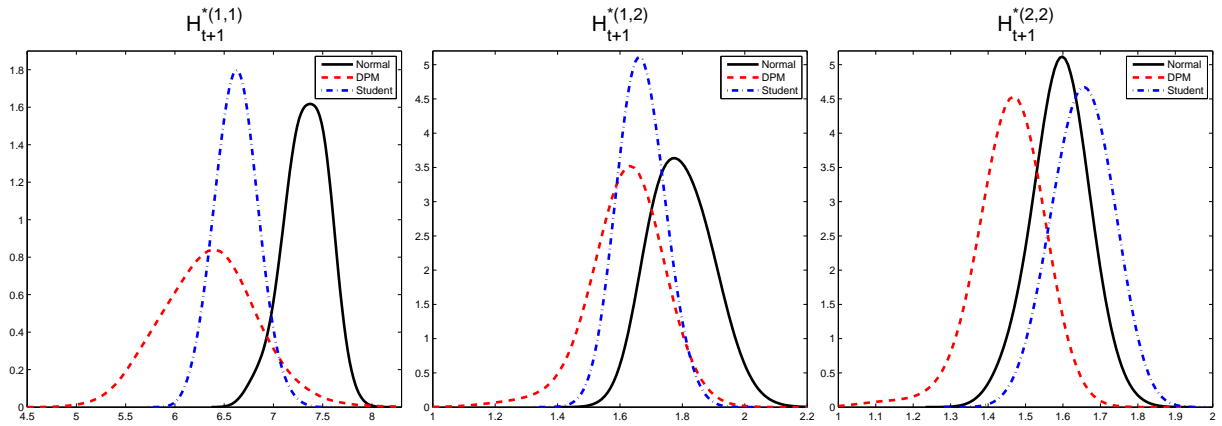**Figure 7.** Posterior Distributions of One-step-ahead Volatilities for the Three Errors Specifications

Table 6. Posterior mean, median and 95% credible intervals for the optimal one-step-ahead portfolio weight, variance and return.

| | Gaussian | | | Student-t | | | DPM | |
| | Mean | 95% CI | | Mean | 95% CI | | Mean | 95% CI |
	Median			Median			Median	
$p^\star_{T+1}$	-0.0401	(-0.0775, -0.0084)		-0.0015	(-0.0245, 0.0204)		-0.0327	(-0.0726, 0.0038)
	-0.0391			-0.0012			-0.0320	
$\sigma^{2\star}_{T+1}$	1.5811	(1.4609, 1.7318)		1.6518	(1.5175, 1.7722)		1.4526	(1.2577, 1.5939)
	1.5771			1.6536			1.4591	
$g^\star_{T+1}$	3.2009	(3.1255, 3.2897)		3.1091	(3.0572, 3.1637)		3.1832	(3.0965, 3.2780)
	3.1986			3.1085			3.1817	

5.2 Portfolio Allocation

Here we are interested in estimating the GMV optimal portfolio of the two real assets, without the short-sale constraint, using the procedure described in Section 3. Firstly, we will make predictions on the optimal one-step-ahead portfolio and then, we will consider all the 233 out-of-sample future observations, adjusting the optimal portfolio weights at each time period. The estimation results for the $T + 1$ period are presented in Table 6. The major difference between the estimates is the width of the credible intervals of the portfolio variance. The same can be observed for the rest of the period, as seen in Figure 10. Therefore, if the investor chooses to be Gaussian or Student, she would underestimate the variance of the variance of her portfolio.

Next, we estimate the optimal portfolio weights for the entire out-of-sample period of 233 observations. Figures 8 and 9 present the dynamics of the estimated portfolio weights and variances for each of the models. It shows that along time the mean portfolio weights are rather similar across all three models. As mentioned before, the differences arise in the thickness of the credible intervals for the portfolio variance, as seen in Figure 10. This allows for a more realistic evaluation of the uncertainty that investor is facing in financial risk management problems.

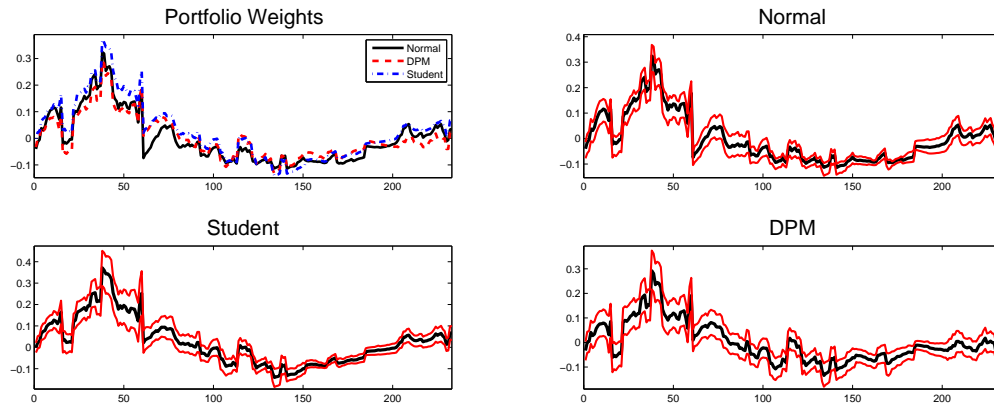**Figure 8.** Dynamics of Portfolio Weights and 95% Credible Intervals.

Figure 9. Dynamics of Portfolio Variance and 95% Credible Intervals.
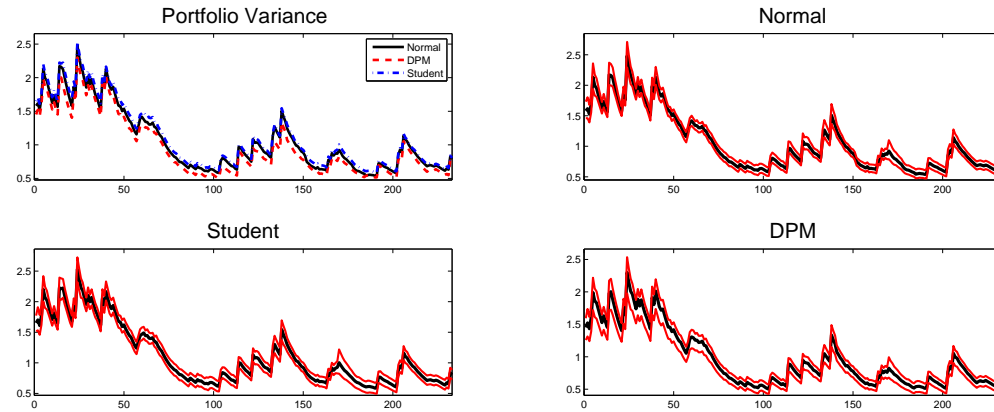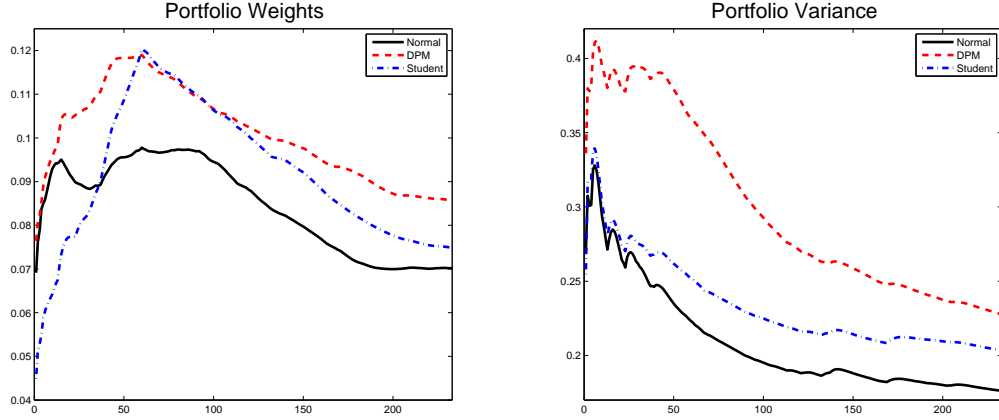
Figure 10. Mean Cumsum of the Width of the 95% Credible Intervals for Portfolio Weights and Variance.

To sum up, these portfolio allocation exercises helped to illustrate the direct consequences of return distribution to the uncertainty of financial decisions. The DPM model permits the investor to perform inference and prediction about the returns and their volatilities without imposing arbitrary restrictions on the data generating process. In the portfolio allocation context, adjusting portfolio weights at each period might lead to high transaction costs, thus the investor will adjust her portfolio only if the expected utility after the adjustment minus the transaction costs is greater than the expected utility without the adjustment. The illustration has shown the differences in error specifications in using real data. We have illustrated how quantification of uncertainty reflects distributional assumptions of the errors.

6.   Conclusions

In this paper we have proposed a Bayesian non-parametric approach for modeling the distribution of multiple returns. We have used an ADCC-GJR-GARCH model to explain the individual volatilities and the time-varying correlations and taking into consideration the asymmetries in individual assets' volatilities, as well as in the correlations. The errors are modeled using a location-scale mixture of infinite Gaussian distributions that has been shown

to allow for a great flexibility in the return distribution in terms of skewness and kurtosis. An MCMC method has been presented for model estimation and prediction. For that, a Dirichlet process mixture (DPM) prior has been given to the infinite mixture of multivariate Gaussian distribution. We have also considered a dynamic portfolio allocation problem, where the time-varying covariance matrix is estimated using a ADCC-GJR-GARCH model. We have presented a short simulation study that illustrates the differences arising from different assumptions for the errors and shows the adaptability of the DPM model. The simulation results suggest that the proposed approach appears to be able to fit adequately several frequently used distributions. Here is the main importance of our approach because in practice one never knows which is the true error distribution. Finally, we have presented an application to return data of Apple Inc. and NASDAQ Industrial that compares the DPM specification with a Gaussian and Student-t distributions. Model comparison via log-predictive likelihood favors the non-parametric approach. Additionally, we have employed the proposed approach to solve a portfolio allocation problem. In the application we have showed that even though the point estimates for optimal portfolio weights are very similar for Gaussian or Student-t, the non-parametric credible intervals for the volatilities are wider. Therefore, the normality or Student assumptions forces the investor to be overconfident about her estimates. The explained methodology and obtained results are not limited to this specific risk management problem and could be expanded into various other topics in applied finance and risk management.

Acknowledgements

References

Antoniak, C. E. (1974). Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics*, *2*, 1152–1174. doi:`10.1214/aos/1176342871`.

Ardia, D., & Hoogerheide, L. F. (2010). Efficient Bayesian Estimation and Combination of GARCH-Type Models. In K. Böcker (Ed.), *Rethinking Risk Measurement and Reporting: Examples and Applications from Finance: Vol II* chapter 1. (pp. 3–30). London: RiskBooks.

Ausín, M. C., & Galeano, P. (2007). Bayesian Estimation of the Gaussian Mixture GARCH Model. *Computational Statistics & Data Analysis*, *51*, 2636–2652. doi:`10.1016/j.csda.2006.01.006`.

Ausín, M. C., Galeano, P., & Ghosh, P. (2014). A semiparametric Bayesian approach to the analysis of financial time series with applications to value at risk estimation. *European Journal of Operational Research*, *232*, 350–358. doi:`10.1016/j.ejor.2013.07.008`.

Avramov, D., & Zhou, G. (2010). Bayesian Portfolio Analysis. *Annual Review of Financial Economics*, *2*, 25–47. doi:`10.1146/annurev-financial-120209-133947`.

Bauwens, L., & Laurent, S. (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics*, *23*, 346–354. doi:`10.1198/073500104000000523`.

Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2006). Multivariate GARCH Models: a Survey. *Journal of Applied Econometrics*, *21*, 79–109. doi:`10.1002/jae.842`.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, *31*, 307 – 327. doi:`10.1016/0304-4076(86)90063-1`.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence. *Journal of Econometrics*, *52*, 5–59. doi:`10.1016/0304-4076(92)90064-X`.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH Models. In R. F. Engle, & D. McFadden (Eds.), *Handbook of Econometrics, Vol. 4* (pp. 2959–3038). Amsterdam: Elsevier Science.

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, *4*, 537–572. doi:`10.1093/jjfinec/nbl005`.

Cecchetti, S. G., Cumby, R. E., & Figlewski, S. (1988). Estimation of the Optimal Futures Hedge. *The Review of Economics and Statistics*, *70*, 623–630. doi:`10.2307/1935825`.

Christodoulakis, G. a., & Satchell, S. E. (2002). Correlated ARCH (CorrARCH): Modelling the Time-varying Conditional Correlation Between Financial Asset Returns. *European Journal of Operational Research*, *139*, 351–370. doi:`10.1016/S0377-2217(01)00361-7`.

Delatola, E.-I., & Griffin, J. (2013). A Bayesian Semiparametric Model for Volatility with a Leverage Effect. *Computational Statistics & Data Analysis*, *60*, 97–110. doi:`10.1016/j.csda.2012.10.023`.

Delatola, E.-I., & Griffin, J. E. (2011). Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility. *Bayesian Analysis*, *6*, 901–926. doi:`10.1214/11-BA632`.

Ding, Z., & Engle, R. F. (2001). Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing. *NYU working paper*, *29*, 157–184.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, *50*, 987. doi:`10.2307/1912773`.

Engle, R. F. (2002a). Dynamic Conditional Correlation: a Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics*, *20*, 339–350. doi:`10.1198/073500102288618487`.

Engle, R. F. (2002b). New Frontiers for ARCH Models. *Journal of Applied Econometrics*, *17*, 425–446. doi:`10.1002/jae.683`.

Escobar, M. D., & West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, *90*, 577–588. doi:`10.2307/2291069`.

Fiorentini, G., Sentana, E., & Calzolari, G. (2003). Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with student t innovations. *Journal of Business and Economic Statistics*, *21*, 532–546. doi:`10.1198/073500103288619232`.

Galeano, P., & Ausín, M. C. (2010). The Gaussian mixture dynamic conditional correlation model: parameter estimation, value at risk calculation, and portfolio selection. *Journal of Business and Economic Statistics*, *28*, 559–571. doi:`10.1198/jbes.2009.07238`.

Giamouridis, D., & Vrontos, I. D. (2007). Hedge Fund Portfolio Construction: a Comparison of Static and Dynamic Approaches. *Journal of Banking & Finance*, *31*, 199–217. doi:`10.1016/j.jbankfin.2006.01.002`.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, *48*, 1779–1801. doi:`10.2307/2329067`.

Greyserman, A., Jones, D. H., & Strawderman, W. E. (2006). Portfolio Selection Using Hierarchical Bayesian Analysis and MCMC Methods. *Journal of Banking & Finance*, *30*, 669–678. doi:`10.1016/j.jbankfin.2005.04.008`.

Hentschel, L. (1995). All in the Family Nesting Symmetric and Asymmetric GARCH Models. *Journal of Financial Economics*, *39*, 71–104. doi:10.1016/0304-405X(94)00821-H.

Jacquier, E., & Polson, N. G. (2013). Asset Allocation in Finance: A Bayesian Perspective. In P. Demian, P. Dellaportas, N. G. Polson, & D. A. Stephen (Eds.), *Bayesian Theory and Applications* chapter 25. (pp. 501–516). Oxford: Oxford University Press. (1st ed.).

Jensen, M. J., & Maheu, J. M. (2010). Bayesian Semiparametric Stochastic Volatility Modeling. *Journal of Econometrics*, *157*, 306–316. doi:10.1016/j.jeconom.2010.01.014.

Jensen, M. J., & Maheu, J. M. (2012). Estimating a Semiparametric Asymmetric Stochastic Volatility Model with a Dirichlet Process Mixture. *Working Paper, University of Toronto*, *453*.

Jensen, M. J., & Maheu, J. M. (2013). Bayesian Semiparametric Multivariate GARCH Modeling. *Journal of Econometrics*, *176*, 3–17. doi:10.1016/j.jeconom.2013.03.009.

Jorion, P. (1986). Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis*, *21*, 279–292. doi:10.2307/2331042.

Kang, L. (2011). Asset Allocation in a Bayesian Copula-GARCH Framework: an Application to the "Passive Funds Versus Active Funds" Problem. *Journal of Asset Management*, *12*, 45–66. doi:10.1057/jam.2010.6.

Kroner, K. F., & Sultan, J. (1993). Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *The Journal of Financial and Quantitative Analysis*, *28*, 535–551. doi:10.2307/2331164.

Markowitz, H. M. (1952). Portfolio Selection. *Journal of Finance*, *7*, 77–91. doi:10.2307/2975974.

Merton, R. C. (1972). An Analytic Derivation of the Efficient Portfolio Frontier. *The Journal of Financial and Quantitative Analysis*, *7*, 1851–1872. doi:10.2307/2329621.

Papaspiliopoulos, O. (2008). A Note on Posterior Sampling from Dirichlet Mixture Models. *Working Paper: Department of Economics, Universitat Pompeu Fabra, Barcelona*, (pp. 1–8).

Papaspiliopoulos, O., & Roberts, G. O. (2008). Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models. *Biometrika*, *95*, 169–186. doi:`10.1093/biomet/asm086`.

Robert, C. P., & Casella, G. (2005). *Monte Carlo Statistical Methods*. (2nd ed.). New York: Springer.

Rossi, E., & Spazzini, F. (2010). Model and Distribution Uncertainty in Multivariate GARCH Estimation: A Monte Carlo Analysis. *Computational Statistics & Data Analysis*, *54*, 2786–2800. doi:`10.1016/j.csda.2009.06.004`.

Rossi, E., & Zucca, C. (2002). Hedging Interest Rate Risk with Multivariate GARCH. *Applied Financial Economics*, *12*, 241–251. doi:`10.1080/09603100110088094`.

Sethuraman, J. (1994). A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, *4*, 639–650.

Silvennoinen, A., & Teräsvirta, T. (2009). Multivariate GARCH Models. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, & T. Mikosch (Eds.), *Handbook of Financial Time Series* (pp. 201–229). Berlin, Heidelberg: Springer.

Teräsvirta, T. (2009). An Introduction to Univariate GARCH Models. In T. G. Andersen, R. A. Davis, J.-P. Kreiß, & T. Mikosch (Eds.), *Handbook of Financial Time Series* (pp. 17–42). Berlin, Heidelberg: Springer.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. (3rd ed.). Hoboken: John Willey & Sons.

Tse, Y., & Tsui, A. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business and Economic Statistics*, *20*, 351–362. doi:10.1198/073500102288618496.

Virbickaite, A., Ausín, M. C., & Galeano, P. (2013). Bayesian Inference Methods for Univariate and Multivariate Garch Models: a Survey. *Journal of Economic Surveys*, *In Press*. doi:10.1111/joes.12046.

Walker, S. G. (2007). Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation*, *36*, 45–54. doi:10.1080/03610910601096262.

Yang, W., & Allen, D. E. (2005). Multivariate GARCH Hedge Ratios and Hedging Effectiveness in Australian Futures Markets. *Accounting and Finance*, *45*, 301–321. doi:10.1111/j.1467-629x.2004.00119.x.

Yilmaz, T. (2011). Improving Portfolio Optimization by DCC and DECO GARCH: Evidence from Istanbul Stock Exchange. *International Journal of Advanced Economics and Business Management*, *1*, 81–92.