



Working Paper 13-10  
Statistics and Econometrics Series 09  
May 2013

Departamento de Estadística  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (34) 91 624-98-49

**A BAYESIAN NON-PARAMETRIC APPROACH TO ASYMMETRIC  
DYNAMIC CONDITIONAL CORRELATION MODEL WITH APPLICATION  
TO PORTFOLIO SELECTION**

Audrone Virbickaite, M. Concepción Ausín and Pedro Galeano

**Abstract**

We use an asymmetric dynamic conditional correlation (ADCC) GJR-GARCH model to estimate the time-varying volatilities of financial returns. 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 flexible location-scale mixture of infinite Gaussian distributions and the inference and estimation is carried out by relying on Bayesian non-parametrics. Finally, we carry out a simulation study to illustrate the flexibility of the new method and present a financial application using Apple and NASDAQ Industrial index data to solve a portfolio allocation problem.

**Keywords:** Asymmetric Dynamic Condition Correlation; Bayesian Non-parametrics; Dirichlet Process Mixtures; Portfolio Allocation.

The first and second authors are grateful for the financial support from MEC grant ECO2011-25706. The third author acknowledges financial support from MEC grant ECO2012-38442. Universidad Carlos III de Madrid, c/ Madrid 126, Getafe (Madrid) 28903, Dpto. Estadística. Contact information: Audrone Virbickaite, tel.: 916249674, fax: 916249848, email: [audrone.virbickaite@uc3m.es](mailto:audrone.virbickaite@uc3m.es). M. Concepción Ausín, tel.: 916245852, fax: 916249848, email: [concepcion.ausin@uc3m.es](mailto:concepcion.ausin@uc3m.es). Pedro Galeano, tel.: 916248901, fax: 916249848, email: [pedro.galeano@uc3m.es](mailto:pedro.galeano@uc3m.es).

# A Bayesian Non-Parametric Approach to Asymmetric Dynamic Conditional Correlation Model with Application to Portfolio Selection\*

Audrone Virbickaite<sup>†</sup>, M. Concepción Ausín<sup>‡</sup> and Pedro Galeano<sup>§</sup>

## Abstract

We use an asymmetric dynamic conditional correlation (ADCC) GJR-GARCH model to estimate the time-varying volatilities of financial returns. 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 flexible location-scale mixture of infinite Gaussian distributions and the estimation and prediction are done by relying on Bayesian non-parametrics. Finally, we carry out a simulation study to illustrate the flexibility of the new method and present a financial application using Apple and NASDAQ Industrial index data to solve a portfolio allocation problem.

**Keywords:** Asymmetric Dynamic Condition Correlation; Bayesian Non-Parametrics; Dirichlet Process Mixtures; Portfolio Allocation.

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

---

\*The first and second authors are grateful for the financial support from MEC grant ECO2011-25706. The third author acknowledges financial support from MEC grant ECO2012-38442.

<sup>†</sup>Universidad Carlos III de Madrid, c/ Madrid 126, Getafe (Madrid) 28903, Dpto. Estadística (UC3M); tel.: 916249674; fax: 916249848; email: audrone.virbickaite@uc3m.es.

<sup>‡</sup>UC3M; tel.: 916245852; fax: 916249848; email: concepcion.ausin@uc3m.es.

<sup>§</sup>UC3M; tel.: 916248901; fax: 916249848; email: pedro.galeano@uc3m.es.

# 1 Introduction

Modeling the dynamics of the assets' returns has been extensively researched for decades and the topic yet remains of great interest, especially in empirical finance setting. ARCH-family models, first introduced by Engle (1982) and then generalized by Bollerslev (1986), without doubt, are the most researched and used in practice to explain time-varying volatilities, see also Bollerslev *et al.* (1992), Bollerslev *et al.* (1994), Engle (2002b), Teräsvirta (2009) and Tsay (2010). When dealing with multivariate returns, one must also take into consideration the mutual dependence between them. In this sense, dynamic conditional correlation models plays an important role. For recent reviews on multivariate GARCH (MGARCH) models, see Bauwens *et al.* (2006), Silvennoinen and Teräsvirta (2009) and Tsay (2010).

In this multivariate setting, correlations may also exhibit some stylized features, such as persistence and asymmetry. The asymmetric behavior of individual returns has been well established in the financial literature, see Hentschel (1995), among others. However, the use of models, explaining asymmetric behavior of covariances, is far less common, even though these effects exist, see Cappiello *et al.* (2006). In this paper, we assume a general MGARCH model which allows for asymmetries not only in individual assets' returns, but also in their correlations. In particular, we consider the Asymmetric Dynamic Conditional Correlation (ADCC) model, proposed by Cappiello *et al.* (2006), with individual GJR-GARCH models, proposed by Glosten *et al.* (1993). This specification provides a much more realistic evaluation of the co-movements of the assets' returns than standard symmetric MGARCH models.

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 Normal or Student-t distributions may be rather restrictive. Alternatively, in this

paper, we propose a Bayesian non-parametric approach for ADCC models avoiding the specification of a particular parametric distribution for the return innovations. More specifically, we consider a Dirichlet Process Mixture (DPM) model, firstly introduced by Antoniak (1974), with a Gaussian base distribution. 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.* (2011), who have applied the DPM models for univariate GJR-GARCH, and Jensen and Maheu (2012), who have used DPM models for the multivariate symmetric DVEC by Ding and Engle (2001).

The Bayesian approach also helps to deal with parameter uncertainty in portfolio decision problems, see e.g. Jorion (1986), Greyserman *et al.* (2006), Avramov and 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 and 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.

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 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.* (2011) to the multivariate framework and the recent work by Jensen and Maheu (2012) to the asymmetric setting. Also, differently from the work of Jensen and Maheu (2012), 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 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 Asymmetric DCC 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) (based on the previous work by Engle, 2002a) to model joint volatilities. This leads to the following final model:

$$r_t = H_t^{1/2} \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{F}_K, \quad (1)$$

$$H_t = D_t R_t D_t, \quad (2)$$

$$D_t^2 = \text{diag}(\omega_i) + [\text{diag}(\alpha_i) + \text{diag}(\text{diag}(\phi_i) I'_{t-1})] \odot r'_{t-1} r_{t-1} + \text{diag}(\beta_i) \odot D_{t-1}^2, \quad (3)$$

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

$$Q_t = S(1 - \kappa - \lambda - \delta/2) + \kappa \times \epsilon'_{t-1} \epsilon_{t-1} + \lambda \times Q_{t-1} + \delta \times \eta'_{t-1} \eta_{t-1}, \quad (5)$$

$$R_t = (\text{diag}(Q_t))^{-1/2} Q_t (\text{diag}(Q_t))^{-1/2}, \quad (6)$$

where  $\mathcal{F}_K$  is an unknown  $K$ -dimensional distribution for which we will assume a DPM prior specified later,  $\text{diag}$  stands for either taking just the diagonal elements from the matrix, or making a diagonal matrix from a vector,  $S$  is a sample correlation matrix of  $\epsilon_t$  and  $\odot$  denotes Hadamard matrix product operator. Individual volatilities are represented in the Equation (1):  $d_{iit}^2 = \omega_i + (\alpha_i + \phi_i I_{i,t-1}) r_{it-1}^2 + \beta d_{iit-1}^2$ , where  $I_{i,t-1}$  is an indicator function  $I_{i,t-1}(r_{it} < 0)$  and  $d_{iit}^2$  are individual asset volatilities, following a GJR-GARCH model with parameters  $\omega_i, \alpha_i, \phi_i, \beta_i > 0$ . To ensure the positivity and stationarity of  $Q_t$ , we impose  $\kappa, \lambda, \delta > 0$  and  $\kappa + \lambda + \delta/2 < 1$ . Let us define  $\Phi = (\omega, \alpha, \beta, \phi, \kappa, \lambda, \delta)$  as the set of parameters describing the dynamics of individual volatilities and correlations.

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 with Gaussian base and some of the most important special cases arising from this model. Using the stick-breaking

representation by Sethuraman (1994), it can be expressed as an infinite location-scale mixture of Normal distributions. Then, the resulting ADCC-DPM model can be written as:

$$r_t = H_t^{1/2} \epsilon_t, \text{ where } f(\epsilon_t | \rho, \mu, \Lambda) = \sum_{i=1}^{\infty} \rho_i \mathcal{N}_K(\epsilon_t | \mu_i, \Lambda_i^{-1}),$$

where the individual and joint dynamics of the model are as seen before. Here  $\mathcal{N}_K$  denotes a  $K$ -variate normal density. Let us denote by  $\Omega = \{\rho_i, \mu_i, \Lambda_i\}_{i=1}^{\infty}$  the infinite-dimensional parameter vector describing the innovation mixture distribution. Here  $\rho_i$  represent the component weights,  $\mu_i$  are the component means and  $\Lambda_i$  are the the precision matrices. Using the stick breaking representation, the weights of the infinite mixture components are reparameterized as follows:  $\rho_1 = v_1$ ,  $\rho_i = (1 - v_1) \dots (1 - v_{i-1})v_i$ , where a Beta prior distribution is assumed for  $v_i \sim \mathcal{B}(1, c)$ , for  $i = 1, 2, \dots$ . It is well known that in DPM models, there is usually important sensitivity to the choice of the concentration parameter,  $c$ . Therefore, we further assume a Gamma hyper-prior distribution,  $c \sim \mathcal{G}(a_0, b_0)$ . As a base distribution, we assume a conjugate Normal-Wishart prior for  $\Lambda$  and  $\mu$ :

$$\begin{aligned} \mu_i, \Lambda_i &\sim \mathcal{NW}(\mu_0, s_0, W, df), \\ \mu_i | \mu_0, s_0, \Lambda &\sim \mathcal{N}_K(\mu | \mu_0, (s_0 \Lambda)^{-1}), \\ \Lambda_i | W, df &\sim \mathcal{W}(\Lambda | W, df), \end{aligned}$$

for  $i = 1, 2, \dots$ , such that  $E[\Lambda] = df \times W^{-1}$  and  $E[\Lambda^{-1}] = (df - (K + 1)/2)^{-1} \times W$ .

Therefore, the complete set of model parameters is denoted by  $\Theta = (\Phi, \Omega)$  and, given the information available up to time  $t - 1$ , denoted by  $r^{t-1} = (r_1, \dots, r_{t-1})$ , the conditional density of the returns can be written as follows:

$$f(r_t | \Theta, r^{t-1}) = \sum_{i=1}^{\infty} \rho_i \mathcal{N}_K \left( H_t^{1/2} \mu_i, H_t^{1/2} \Lambda_i^{-1} (H_t^{1/2})' \right), \quad (7)$$

with conditional moments given by

$$E [r_t | \Theta, r^{t-1}] = H_t^{1/2} \left( \sum_{i=1}^{\infty} \rho_i \mu_i \right)$$

and

$$Cov [r_t | \Theta, r^{t-1}] = H_t^{1/2} Cov [\epsilon_t] (H_t^{1/2})',$$

where

$$Cov [\epsilon_t | \Omega] = \sum_{i=1}^{\infty} \rho_i (\Lambda_i^{-1} + \mu_i (\mu_i)') - \left( \sum_{i=1}^{\infty} \rho_i \mu_i \right) \left( \sum_{i=1}^{\infty} \rho_i \mu_i \right)'$$

It is important to notice that this full unrestricted model induces the GARCH-in-Mean effects, since the conditional mean of the returns is not restricted to be zero. 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 we have 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  $\mu_i = 0$ , for  $i = 1, 2, \dots$ . If we further assume that the precision matrices are all diagonal,  $\Lambda_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{iK})$ , this will lead to uncorrelated innovations. Finally, we could in addition assume that the diagonal elements of each precision matrix are all equal,  $\Lambda_i = \lambda_i I_K$ . In this paper we will use the full version of the DPM model to illustrate the adaptability of it. However, it is straightforward how to adapt the model to these particular cases in order to simplify the problem of many assets.

## 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 GJR-ADCC-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 and 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 and West (1995) produces 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  $(\epsilon, u)$  given  $\Omega = (\rho, \mu, \Lambda)$  is given by

$$f(\epsilon_t, u_t | \Omega) = \sum_{i=1}^{\infty} \mathbf{1}(u_t < \rho_i) \mathcal{N}_K(\epsilon_t | \mu_i, \Lambda_i^{-1}). \quad (8)$$

Let  $A_\rho(u_t) = \{i : \rho_i > u_t\}$  be a set of size  $N_{u_t}$ , which is finite for all  $u_t > 0$ . Then the joint density of  $(\epsilon_t, u_t)$  in (8) can be equivalently written as  $f(\epsilon_t, u_t | \Omega) = \sum_{i \in A_\rho(u_t)} \mathcal{N}_K(\epsilon_t | \mu_i, \Lambda_i^{-1})$ . Integrating over  $u_t$  gives us the density of infinite mixture of distributions. Finally, given  $u_t$ , the number of mixture components is finite. In order to simplify the likelihood, we also need to introduce further indicator latent variable  $z_t$ , which indicates the mixture component that  $\epsilon_t$  comes from:  $f(\epsilon_t, z_t = j, u_t | \Omega) = \mathcal{N}_K(\epsilon_t | \mu_j, \Lambda_j^{-1}) \mathbf{1}(j \in A_\rho(u_t))$ . The log-likelihood of  $\Theta$ , given the latent variables  $u_t$  and  $z_t$  looks as follows:

$$\log L(\Theta | u_t, z_t) = -\frac{1}{2} \sum_{t=1}^T (k \log(2\pi) + \log |H_t^*| + (r_t - \mu_t^*) H_t^{*-1} (r_t - \mu_t^*)'), \quad (9)$$

where  $\mu_t^*$  is the conditional mean vector and  $H_t^*$  is the conditional covariance matrix:

$$\begin{aligned}\mu_t^* &= \mathbb{E} [r_t | r^{t-1}, z_t] = H_t^{1/2} \mu_{z_t}, \\ H_t^* &= \text{Cov} [r_t | r^{t-1}, z_t] = H_t^{1/2} \Lambda_{z_t}^{-1} H_t^{1/2}.\end{aligned}\tag{10}$$

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

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

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

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

In the **second step**, we sample the weights of the components  $v_j$  for  $j = 1, \dots, z^*$ , where the prior for  $v \sim \mathcal{B}(1, c)$  and, given the data and  $z$ :

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

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

At the **third step**, we update  $u_t \sim \mathcal{U}(0, \rho_{z_t})$ , for  $t = 1, \dots, T$ .

In the **fourth step**, we sample all the values of  $\rho_j$  that are larger than  $u_t$ . As Walker (2007) showed we need to find the smallest  $j^*$  such that  $\sum_{j=1}^{j^*} \rho_j > u^*$  and then update  $v_j$  and  $\rho_j$  for  $j = z^* + 1, \dots, j^*$ , where  $u^* = \min(u_1, \dots, u_T)$ .

Next, update  $\mu$  and  $\Lambda$ , whose posterior distribution is independent of  $(\rho, u_t)$ :

$$\begin{aligned} p(\mu, \Lambda | \epsilon^t) &= \mathcal{N}_K(\mu | \mu_n, ((s_0 + n)\Lambda)^{-1}) \mathcal{W}(\Lambda | W_n, df + n), \\ \mu_n &= \frac{s_0\mu_0 + n\bar{\epsilon}}{s_0 + n}, \\ W_n &= W^{-1} + S + \frac{s_0n}{s_0 + n}(\mu_0 - \bar{\epsilon})(\mu_0 - \bar{\epsilon})', \\ S &= \sum_{i=1}^n (\epsilon_i - \bar{\epsilon})(\epsilon_i - \bar{\epsilon}). \end{aligned}$$

Note that this approach is different from the one described in Jensen and Maheu (2012), because they assume independent prior distributions for  $\mu$  and  $\Lambda$ , and therefore, they need to include some Gibbs steps to sample from the conditional posterior. In our case this is not necessary since we have assumed a Normal-Wishart prior for  $(\mu, \Lambda)$ .

In the **sixth step**, we update to which component the observations belong to by using the following (as seen in Walker (2007)):

$$\text{Probability}(z_t = j | \dots) \propto \mathbf{1}(j \in A_\rho(u_t)) \mathcal{N}_K(\epsilon_t | \mu_j, \Lambda_j^{-1}),$$

where  $A_\rho(u_t) = \{j : \rho_j > u_t\}$ , which is not empty.

The rest of the steps of the algorithm concern updating the parameters of the GJR-ADCC model. We use the Random Walk Metropolis Hasting (RWMH), following the same procedure as in Jensen and Maheu (2012). For set of parameters  $\Phi$  a candidate value  $\tilde{\Phi}$  is generated from a  $P$ -variate Normal distribution with mean equal to the previous value of the parameter, where  $P$  is the number of parameters in  $\Phi$  as follows:

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

The probability of accepting a proposed value  $\tilde{\Phi}$ , given the current value  $\Phi$ , is

$\alpha(\Phi, \tilde{\Phi}) = \min \left\{ 1, \frac{\prod_{t=1}^T l(r_t|\Phi)}{\prod_{t=1}^T \tilde{l}(r_t|\tilde{\Phi})} \right\}$ , where the likelihood used is as in (9), see e.g. Robert and Casella (2004). 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%. In our applications 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:

$$f(r_{t+1}|r^t) = \int f(r_{t+1}|\Theta, r^t) f(\Theta|r^t) d\Theta, \quad (11)$$

where  $f(r_{t+1}|\Theta, r^t)$  as specified in (7). Although this integral is not analytically tractable, we can approximate it using the MCMC output. For this, we make use of the procedure described in Walker (2007). At the  $m$ -th iteration of the MCMC algorithm, there are weights  $\rho_j^{(m)}$  and corresponding pairs of means  $\mu_j^{(m)}$  and precision matrices  $\Lambda_j^{(m)}$ , for  $j = 1, \dots, j^{*(m)}$ , then,

- i. We sample a random variable  $r^{(m)} \sim \mathcal{U}(0, 1)$ .
- ii. Take such  $\rho_s^{(m)}$  for which  $\sum_{j=1}^{s-1} \rho_{j-1}^{(m)} < r^{(m)} < \sum_{j=1}^s \rho_j^{(m)}$  and the corresponding pair  $(\mu_s, \Lambda_s)^{(m)}$ .
- iii. If we need more weights (i.e.  $\sum_{j=1}^{j^{*(m)}} \rho_j^{(m)} < r^{(m)}$ ), we can sample additional  $\rho_j^{(m)}$  as before<sup>1</sup>, and  $(\mu_j, \Lambda_j)^{(m)}$  from the Normal - Wishart prior.

---

<sup>1</sup> $\rho_1 = v_1, \rho_j = (1 - v_1) \dots (1 - v_{j-1}) v_j, v_j \sim \mathcal{B}(1, c)$

Therefore, the one-step-ahead predictive density of the returns can be approximated by:

$$f(r_{t+1}|r^t) \simeq \frac{1}{M} \sum_{m=1}^M f\left(r_{t+1} | \mu_s^{(m)} H_{t+1}^{1/2,(m)}, H_{t+1}^{1/2,(m)} \Lambda_s^{-1,(m)} (H_t^{1/2,(m)})'\right), \quad (12)$$

where  $(\mu_s, \Lambda_s)^{(m)}$  is the pair of mean and precision matrix simulated at each MCMC iteration as explained before and  $H_{t+1}^{(m)}$  is a one-step-ahead volatility at the  $m$ -th MCMC iteration and  $M$  is the length of the MCMC chain.

Note that Jensen and Maheu (2012) propose a different approach, which is to sample two components instead of one at each MCMC iteration, but we have not found significant differences in this. Walker (2007), on the other hand, proposes sampling only one observation from each component, thus obtaining an actual sample from a predictive density. This approach has the disadvantage that it does not provide a smooth posterior density. However, it maybe useful in order to easily estimate any quantity of interest, such as the predictive mean or variance, using sample approximations.

### 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 and Polson (2012), 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. The standard approach is to assume that the returns  $r_t$  have a constant covariance matrix,  $\Sigma = \text{Cov}[r_t]$ . In this case, the optimization problem and the optimal solution are given by:

$$p^* = \arg \min_p \text{Var}[r_t^P] : p'1_K = 1,$$

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

where  $p$  is the weight vector,  $1_K$  is a  $K$ -vector of ones and  $r_t^P = p'r_t$  is a vector of portfolio returns. Note that, if we choose to impose the short sale constraint, i.e.,  $p_i \geq 0, \forall i = 1, \dots, K$ , the problem cannot be solved analytically anymore and it requires numerical optimization techniques.

However, 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, see Yilmaz (2011). Giamouridis and Vrontos (2007) find that portfolios, constructed under a dynamic approach, have lower average risk and higher out-of-sample risk-adjusted realized return. 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. They show that the use of GARCH-type models leads to the overall portfolio risk reduction, see Rossi and Zucca (2002), Kroner and

Sultan (1993) and Yang and Allen (2004), among others.

To solve the portfolio allocation problem in our case, instead of  $\Sigma = \text{Cov}[r_t]$  we use estimated one-step-ahead conditional covariance matrix for the assets returns  $\text{Cov}[r_{t+1}|r^t] = H_{t+1}^*$ , which is adjusted continuously on the basis of available information up to time  $t$ ,  $r^t$ . Therefore, we are able to obtain optimal portfolio weights for each period as follows:

$$p_{t+1}^*|r^t = \frac{H_{t+1}^{*-1} \mathbf{1}_K}{\mathbf{1}'_K H_{t+1}^{*-1} \mathbf{1}_K}. \quad (13)$$

Using the MCMC output, we can obtain samples from the entire posterior distribution of optimal portfolio weights,  $f(p_{t+1}^*|r^t)$ . 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:

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

where  $\left\{ p_{t+1}^{*(m)} \right\}_{m=1}^M$  is a posterior sample of optimal portfolio weights obtained from (13) for each value of one-step-ahead conditional covariance matrix of the returns,  $\left\{ H_{t+1}^{*(m)} \right\}_{m=1}^M$ , in the MCMC sample. In other words, since we have assembled  $M$  one-step-ahead volatility matrices and mean vectors, we can solve the portfolio allocation problem  $M$  times. Similarly, we can approximate the posterior median of  $p_{t+1}^*$  and credible intervals by using the quantiles of the sample of optimal portfolio weights. In this manner, we are able to obtain a sample that approximates the posterior distribution of the optimal portfolio variance and optimal portfolio return:

$$p(\sigma_{t+1,P}^2|r^t) \sim \left\{ (\sigma_{t+1,P}^2)^{(m)} \right\}_{m=1}^M = \left\{ (p_{t+1}^{*'} H_{t+1}^* p_{t+1}^*)^{(m)} \right\}_{m=1}^M,$$

$$p(r_{t+1,P}|r^t) \sim \left\{ (r_{t+1,P})^{(m)} \right\}_{m=1}^M = \left\{ (p_{t+1}^* r'_{t+1})^{(m)} \right\}_{m=1}^M.$$

## 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 GJR-ADCC model introduced Section 2. We have generated three bivariate time series of 3000 observations from a GJR-ADCC model with the following innovation distributions:

- i. Gaussian  $\mathcal{N}(0, I_2)$ ,
- ii. Student-t  $\mathcal{T}(I_2, \nu = 8)$ ,
- iii. Mixture of two bivariate Normals  $0.9\mathcal{N}(0, \sigma_1^2 = 0.8, \sigma_{12} = 0.0849, \sigma_2^2 = 0.9) + 0.1\mathcal{N}(0, \sigma_1^2 = 2.8, \sigma_{12} = -0.7637, \sigma_2^2 = 1.9)$ .

Note that, in the third case, we have chosen larger variances for the second mixture component to allow for the presence of extreme returns. Then, we estimate all three data sets using the proposed GJR-ADCC-DPM model assuming uninformative uniform priors restricted to the stationary region for  $\Phi$  and setting  $\mu_0 = 0_2$ ,  $s_0 = 0.1$ ,  $df = 5$ ,  $W = I_2/5$ ,  $a_0 = 10$  and  $b_0 = 10$ . The MCMC algorithm is run for 10000 burn-in plus 40000 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. Figures 1 and 2 present estimation results for the three models. Figure 1 draws the average number of non-empty clusters. For normal data DPM model estimates very few non-empty components, 1.2 on average, where there is one dominant weight  $\simeq 1$ . As for the Student-t data, the DPM model estimates many components, around 14 and there are a lot of equal weights. Finally, for the mixture data, the DPM model can identify very well the underlying mixture weights  $\rho_1 = 0.9$  and  $\rho_2 = 0.1$  and the average number of components around 2.8. Figure 2 presents the contour plots, that compare true predictive densities of returns  $f(r_{t+1}|r^t)$  with the estimated ones, which were obtained by sampling one pair of means and precision matrices  $(\mu, \Lambda)^{(m)}$  at each MCMC step. As we can see,

the estimation results are very precise compared to the true contour of one-step-ahead returns. The 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 infinite mixture model is a flexible tool that is able to adjust to whatever distribution the data comes from.

**Figure 1 goes here**

**Figure 2 goes here**

Next, top part of Figure 3 presents the marginal densities for one-step-ahead errors,  $f(\epsilon_{t+1}|r^t)$ , obtained by sampling one observation at each MCMC iteration as in Walker (2007). The Student-t data predicts fatter right and left tails, consequently, allowing for more extreme observations, which in turn increase the volatility. The effect of how fat tails increase volatility can be seen in the middle row of Figure 3, where the kernel smoothing densities of the elements of the sampled error covariance matrices  $\{(\Lambda_{t+1}^{-1})^{(m)}\}_{m=1}^M$  are presented. The Gaussian error data does not allow for extreme observations, therefore does not allow for large variances. However, for the Student-t and mixture data, we observe a fat right tail caused by the more extreme returns. And finally, the bottom row of Figure 3 presents the densities of the sampled volatilities of the returns  $\{(H_{t+1}^*)^{(m)}\}_{m=1}^M$ . Observe that in this case not only the one-step-ahead variances but also the covariances are right skewed for Student-t and mixture data due to the asymmetric correlation effect of the model. This brief simulation study helps to understand and explain the asymmetric posterior distributions that appear in the real data application in the next section.

**Figure 3 goes here**

## 5 Real Data and Results

In this section, we illustrate the proposed procedure using real data set, and solve portfolio allocation problem as described in Section 3.

## 5.1 Estimation

For the illustration we use 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, daily prices are transformed into daily logarithmic returns (in %), resulting in 3098 observations. Table 1 provides the basic descriptive statistics, and Figure 4 illustrates the dynamics of returns.

**Table 1.** 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	

**Figure 4 goes here**

As expected, the Apple Inc. has higher overall variance because of the higher mean return. Both returns do not exhibit any evidence of auto-regressive 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. Data was estimated assuming Gaussian, Student-t and DPM errors.

The estimation results of the ADCC model are reported in the Table 2. As we can see from the table, the constant volatility parameter for the first series is wildly overestimated. The assumption of Gaussian errors does not allow for fat tails, therefore, all the volatility is summed into the  $\omega_1$ . Same happens with the asymmetric volatility parameters  $\phi_1$  and  $\phi_2$  for Gaussian errors. This problem does not appear for the Student-t and DPM models, which produce more or less similar estimates. The posterior mean of the degrees of freedom in the Student-t is around 7. In the DPM model, the average number of non-empty clusters is  $z^* = 7.2$ .

Table 2 also presents the posterior mean and 95% credible interval for the parameter  $A = c/(1 + c)$ . This transformation of the concentration parameter is suggested by Jensen and Maheu (2012) to provide an intuition of the probability of having infinite different clusters in the mixture. Note that this happens as  $c$  goes to infinity which corresponds to the Student-t distribution. Also, Figure 5 draws the histogram of the posterior distribution for  $A$ . Note that the posterior probability that  $A$  is larger than 0.8 is very small, which suggest the better adequacy of the DPM model when compared with the Student-t specification. Note that, different to Jensen and Maheu (2012), we have previously defined a Gamma prior on  $c$  instead of a uniform prior on  $A$ , but we have not observed important prior sensitivity to this choice.

**Figure 5 goes here**

**Table 2.** Estimation Results for Apple Inc. (1) and NASDAQ Ind. (2) Returns, 40,000 iterations plus 10,000 burn-in

	Gaussian		Student-t		DPM	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
$\omega_1$	0.2653	(0.1603, 0.3942)	0.1071	(0.0619, 0.1659)	0.1492	(0.0891, 0.2015)
$\omega_2$	0.0285	(0.0203, 0.0395)	0.0192	(0.0125, 0.0269)	0.0094	(0.0054, 0.0151)
$\alpha_1$	0.0894	(0.0651, 0.1244)	0.0403	(0.0279, 0.0534)	0.0696	(0.0446, 0.0993)
$\alpha_2$	0.0126	(0.0020, 0.0270)	0.0109	(0.0012, 0.0245)	0.0056	(0.0006, 0.0137)
$\beta_1$	0.8430	(0.8039, 0.8740)	0.8975	(0.8730, 0.9204)	0.8885	(0.8720, 0.9018)
$\beta_2$	0.9237	(0.9052, 0.9375)	0.9281	(0.9121, 0.9408)	0.9214	(0.9041, 0.9357)
$\phi_1$	0.1197	(0.0622, 0.1627)	0.0409	(0.0183, 0.0683)	0.0585	(0.0323, 0.0936)
$\phi_2$	0.1050	(0.0796, 0.1312)	0.0739	(0.0520, 0.0912)	0.0397	(0.0235, 0.0612)
$\kappa$	0.0093	(0.0030, 0.0266)	0.0075	(0.0020, 0.0153)	0.0252	(0.0070, 0.0360)
$\lambda$	0.9828	(0.9415, 0.9936)	0.9711	(0.9477, 0.9858)	0.8503	(0.7696, 0.9196)
$\delta$	0.0061	(0.0008, 0.0213)	0.0220	(0.0094, 0.0392)	0.0191	(0.0019, 0.0487)
$\nu$			7.1879	(7.0529, 7.2862)		
$A$					0.5177	(0.3770, 0.6700)
$z^*$					7.2053	(4.0000, 12.0000)

Table 3 present the one-step-ahead means, medians and confidence intervals for the volatilities of the returns, also the main statistics for the degrees of freedom

parameter for the Student distribution and the adjusted one-step-ahead mean for the DPM model  $\mu^* = \mu H_{t+1}^{1/2}$ , as in Equation (10).

**Table 3.** One-Step-Ahead Volatilities for  $r_{t+1}$

	Gaussian		Student-t		DPM	
	Mean Median	95% CI	Mean Median	95% CI	Mean Median	95% CI
$H_{t+1}^{*(1,1)}$	7.3315 7.3419	(6.8461, 7.6932)	4.7799 4.7804	(4.4992, 5.0781)	6.4174 5.9193	(1.7166, 23.5922)
$H_{t+1}^{*(1,2)}$	1.7954 1.7890	(1.6379, 1.9831)	1.2041 1.2017	(1.1263, 1.2973)	1.6689 1.8879	(0.3975, 2.4566)
$H_{t+1}^{*(2,2)}$	1.5932 1.5954	(1.4534, 1.7163)	1.1926 1.1932	(1.0958, 1.2813)	1.4657 1.6492	(0.4700, 2.0146)
$\nu$			7.1879 7.1933	(7.0529, 7.2865)		
$\mu_{t+1}^{*(1)}$	0		0		0.1554 0.1099	(-0.2585, 1.5874)
$\mu_{t+1}^{*(2)}$	0		0		0.0142 -0.0353	(-0.4699, 0.4859)

The credible intervals for the DPM model are much wider, especially for the marginal one-step-ahead volatility of the first series. This is because it allows for some very volatile mixture components, due to the atypical data point. This provides a more realistic evaluation of risk for an agent.

Also, in the DPM model, the posterior distributions for the volatilities are not symmetric. The Apple Inc. one-step-ahead volatility is positively skewed, meanwhile the covariance and the NASDAQ Ind. are negatively skewed. This indicates the flexibility of the DPM model, that it allows for different thickness in tails, as opposed to the Student-t errors, where the degrees of freedom parameters governs the tail thickness in all directions, resulting into symmetric posterior distributions for the volatilities. In order to explore the differences in tails in more detail, we compare predictive one-step-ahead density contours for the three models. Figure 6 draws bivariate contours for the one-step-ahead returns  $r_{t+1}$  and Figure 7 presents the marginal log predictive densities for the returns of Normal, Student-t and DPM

models. Observe that both Normal and Student-t models lead to symmetric predictive densities, although for the Student case it exhibits fatter tails. On the contrary, the DPM model, which allows for different variances and non zero means, predicts an asymmetric multimodal density with fatter tails. Also, as seen from the log marginal one-step-ahead densities, 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 is not so big. Normal model in both cases cannot capture the high kurtosis.

**Figure 6 goes here**

**Figure 7 goes here**

Finally, following Jensen and Maheu (2012), we compare the three estimated models using predictive likelihoods based on a small set of out-of-sample observations,  $\{t + 1, \dots, t + k\}$ . For each new observation, we calculate the predictive likelihood as an average over the MCMC iterations given by:

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

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

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

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. Note that, differently to Jensen and Maheu (2012), 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.** Cumulative log-Predictive Likelihoods for the DPM, Student and Normal Error Models

Model	$\log p(r_{t+1}, \dots, r_{t+k})$
DPM	-746.9376
Student	-749.2850
Normal	-775.8299
# of out-of-sample obs.	k=233

## 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 5.

**Table 5.**  $t + 1$  Portfolio Weights, Variances and Returns

	Gaussian		Student-t		DPM	
	Mean Median	95% CI	Mean Median	95% CI	Mean Median	95% CI
$p_{t+1}^*$	-0.0401 -0.0391	(-0.0775, -0.0084)	-0.0015 -0.0012	(-0.0245, 0.0204)	-0.0500 -0.0560	(-0.1298, 0.0665)
$\sigma_{t+1,P}^2$	1.5811 1.5771	(1.4609, 1.7318)	1.1921 1.1929	(1.0951, 1.2808)	1.4472 1.6309	(0.4560, 1.9967)
$r_{t+1,P}$	3.2009 3.1986	(3.1255, 3.2897)	3.1091 3.1085	(3.0572, 3.1637)	3.2244 3.2385	(2.9478, 3.4138)

The major difference between the estimates is the asymmetry in the portfolio variance in the DPM model. This is caused by the asymmetric covariance, which in turn depends on the asymmetric returns. Therefore, even though the Student errors can incorporate high kurtosis, it cannot capture the skewness in the financial returns. DPM model is more flexible in that sense, because it models errors as a

location-scale mixture. Figure 8 shows the kernel smoothing densities of the one-step-ahead portfolio weights, variances and returns, where we can clearly see the differences in the asymmetries of posterior distributions, especially in the overall portfolio variance.

**Figure 8 goes here**

Next, we estimate the optimal portfolio weights for the entire out-of-sample period of  $T = 233$ . Figures 9, 10 and 11 present the dynamics of the estimated portfolio weights, optimal portfolio variances and optimal portfolio returns for each of the models. In these graphs we can notice again the asymmetric distribution of the volatilities in the DPM model, which is caused by allowing for an asymmetric distribution of the returns. Also, these figures show that along time the mean portfolio weights are rather similar across all three models. The differences arise in the thickness and asymmetry of the credible intervals. DPM model presents asymmetric credible intervals for the volatility of the portfolio, and the asymmetry is more pronounced in volatile periods versus the calm periods. This allows for a more realistic evaluation of the uncertainty that investor is facing in financial risk management problems.

**Figure 9 goes here**

**Figure 10 goes here**

**Figure 11 goes here**

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 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 Conclusion

In this paper we have considered dynamic portfolio allocation problem, where the time-varying covariance matrix was estimated using a GJR-ADCC model, that captures asymmetric volatilities and correlations. For the error term we have used a flexible infinite location-scale mixture of Gaussian distributions, which was handled using a Bayesian non-parametric approach. 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.

We have employed the proposed approach to solve the portfolio allocation problem using the return data of Apple Inc. and NASDAQ Industrial. In the application we have showed that even though the point estimates for optimal portfolio weights are very similar for Gaussian, Student and infinite mixture models, the non-parametric credible intervals are wider and asymmetric. Therefore, the normality assumptions forces the investor to be overconfident about her estimates. Moreover, the non-parametric model allowed for some one-step-ahead volatilities come from a very volatile component, thus making the posterior distribution of covariance matrix asymmetric.

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.

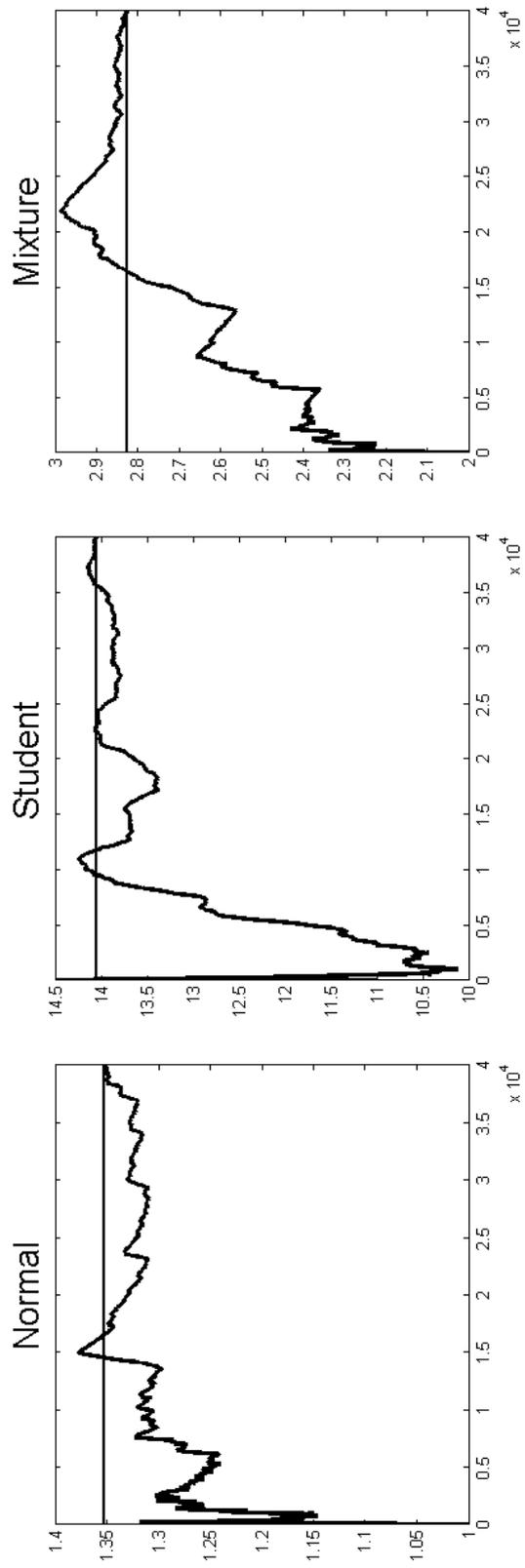
## References

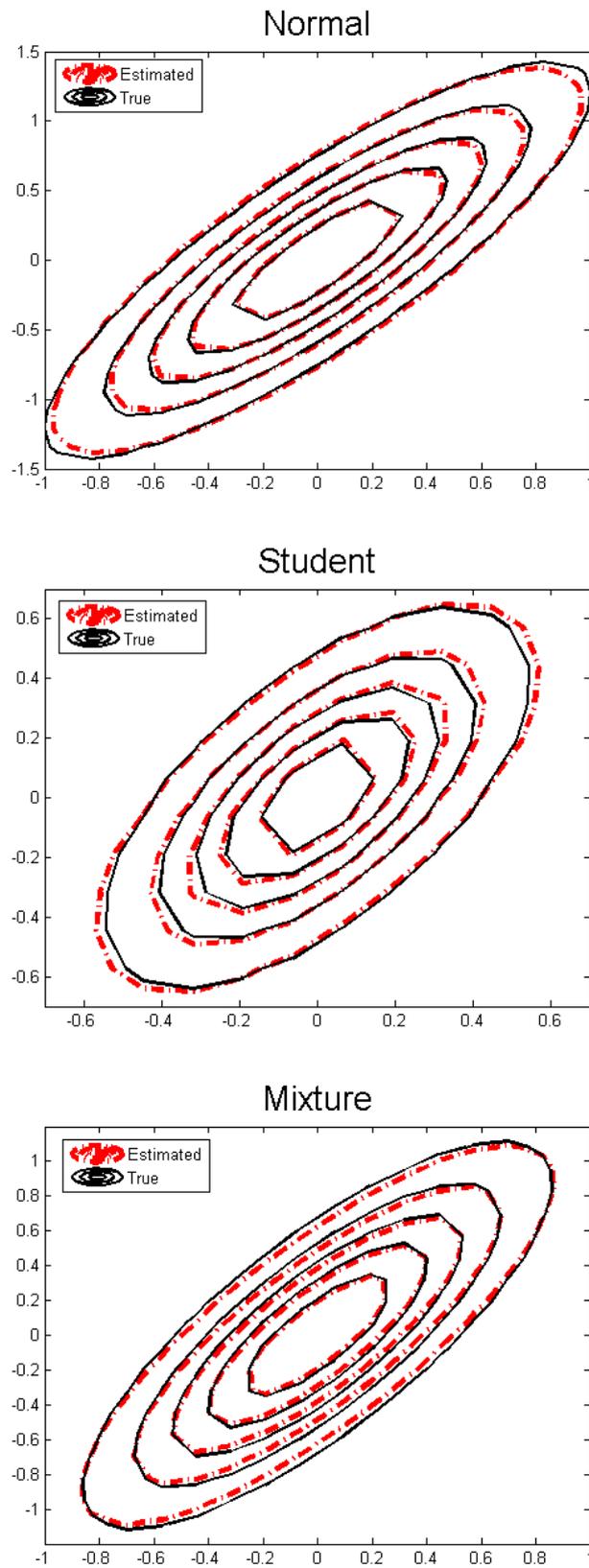
- Antoniak CE. 1974. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics* **2**: 1152–1174.
- Ardia D, Hoogerheide LF. 2010. Efficient Bayesian Estimation and Combination of GARCH-Type Models. In *Rethinking Risk Measurement and Reporting: Examples and Applications from Finance: Vol II*, Bocker K (ed), chap 1. RiskBooks: London.
- Ausín MC, Galeano P, Ghosh P. 2011. A Semiparametric Bayesian Approach to GARCH-Type Models with Application to Value at Risk Estimation. *Working Paper* : 1–33
- Avramov D, Zhou G. 2010. Bayesian Portfolio Analysis. *Annual Review of Financial Economics* **2**: 25–47. ISSN 1941-1367. DOI: 10.1146/annurev-financial-120209-133947
- Bauwens L, Laurent S, Rombouts JVK. 2006. Multivariate GARCH Models: a Survey. *Journal of Applied Econometrics* **21**: 79–109. ISSN 0883-7252. DOI: 10.1002/jae.842
- Bollerslev T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**: 307 – 327.
- Bollerslev T, Chou RY, Kroner KF. 1992. ARCH Modeling in Finance: a Review of the Theory and Empirical Evidence. *Journal of Econometrics* **52**: 5–59.
- Bollerslev T, Engle RF, Nelson DB. 1994. ARCH Models. In *Handbook of Econometrics, Vol. 4*, Engle RF, McFadden D (eds), November. Elsevier: Amsterdam. ISBN 978-0-444-88766-5.
- Cappiello L, Engle RF, Sheppard K. 2006. Asymmetric Dynamics in the Correlations

- of Global Equity and Bond Returns. *Journal of Financial Econometrics* **4**: 537–572.
- Cecchetti SG, Cumby RE, Figlewski S. 1988. Estimation of the Optimal Futures Hedge. *The Review of Economics and Statistics* **70**: 623–630.
- Ding Z, Engle RF. 2001. Large Scale Conditional Covariance Matrix Modeling , Estimation and Testing. *Academia Economic Papers* **29**: 157–184.
- Engle RF. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **50**: 987.
- Engle RF. 2002a. Dynamic Conditional Correlation: a Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business and Economic Statistics* **20**: 339–350. ISSN 0735-0015. DOI: 10.1198/073500102288618487
- Engle RF. 2002b. New Frontiers for ARCH Models. *Journal of Applied Econometrics* **17**: 425–446. ISSN 0883-7252. DOI: 10.1002/jae.683
- Escobar MD, West M. 1995. Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* **90**: 577–588.
- Giamouridis D, Vrontos ID. 2007. Hedge Fund Portfolio Construction: a Comparison of Static and Dynamic Approaches. *Journal of Banking & Finance* **31**: 199–217. ISSN 03784266. DOI: 10.1016/j.jbankfin.2006.01.002
- Glosten LR, Jagannathan R, Runkle DE. 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.
- Greyserman A, Jones DH, Strawderman WE. 2006. Portfolio Selection Using Hierarchical Bayesian Analysis and MCMC Methods. *Journal of Banking & Finance* **30**: 669–678. ISSN 03784266. 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. ISSN 0304405X. DOI: 10.1016/0304-405X(94)00821-H
- Jacquier E, Polson NG. 2012. Asset Allocation in Finance: a Bayesian Perspective. In *Forthcoming in Hierarchical models and MCMC: a Tribute to Adrian Smith*, Damien, Dellaportas, Polson, Stephen (eds). Oxford University Press.
- Jensen MJ, Maheu JM. 2012. Bayesian Semiparametric Multivariate GARCH Modeling. *Working Paper, Federal Reserve Bank of Atlanta* : 1–37
- Jorion P. 1986. Bayes-Stein Estimation for Portfolio Analysis. *The Journal of Financial and Quantitative Analysis* **21**: 279–292.
- 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. ISSN 1470-8272. DOI: 10.1057/jam.2010.6
- Kroner KF, Sultan J. 1993. Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *The Journal of Financial and Quantitative Analysis* **28**: 535–551.
- Markowitz HM. 1952. Portfolio Selection. *Journal of Finance* **7**: 77–91.
- Merton RC. 1972. An Analytic Derivation of the Efficient Portfolio Frontier. *The Journal of Financial and Quantitative Analysis* **7**: 1851–1872.
- Papaspiliopoulos O. 2008. A Note on Posterior Sampling from Dirichlet Mixture Models. *Working Paper* : 1–8
- Papaspiliopoulos O, Roberts GO. 2008. Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models. *Biometrika* **95**: 169–186. ISSN 0006-3444. DOI: 10.1093/biomet/asm086

- Robert CP, Casella G. 2004. *Monte Carlo Statistical Methods*. Springer Texts in Statistics, 2nd edn. ISBN 978-0-387-21239-5.
- Rossi E, Zucca C. 2002. Hedging Interest Rate Risk with Multivariate GARCH. *Applied Financial Economics* **12**: 241–251.
- Sethuraman J. 1994. A Constructive Definition of Dirichlet Priors. *Statistica Sinica* **4**: 639–650.
- Silvennoinen A, Teräsvirta T. 2009. Multivariate GARCH Models. In *Handbook of Financial Time Series*, Mikosch T, KreißJP, Davis RA, Andersen TG (eds). Springer Berlin Heidelberg: Berlin, Heidelberg. ISBN 978-3-540-71296-1.
- Teräsvirta T. 2009. An Introduction to Univariate GARCH Models. In *Handbook of Financial Time Series*, Mikosch T, KreißJP, Davis RA, Andersen TG (eds). Springer Berlin Heidelberg: Berlin, Heidelberg. ISBN 978-3-540-71296-1.
- Tsay RS. 2010. *Analysis of Financial Time Series*. John Wiley & Sons, Inc., 3rd edn. ISBN 9780471690740.
- Walker SG. 2007. Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation* **36**: 45–54. ISSN 0361-0918. DOI: 10.1080/03610910601096262
- Yang W, Allen DE. 2004. Multivariate GARCH Hedge Ratios and Hedging Effectiveness in Australian Futures Markets. *Accounting and Finance* **45**: 301–321. ISSN 0810-5391. 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.

**Figure 1.** The Average Number of Non-Empty Clusters

**Figure 2.** True and Estimated Contours

**Figure 3.** Kernel Smoothing Densities of One-Step-Ahead Errors, their Volatilities and the Volatilities of the Returns

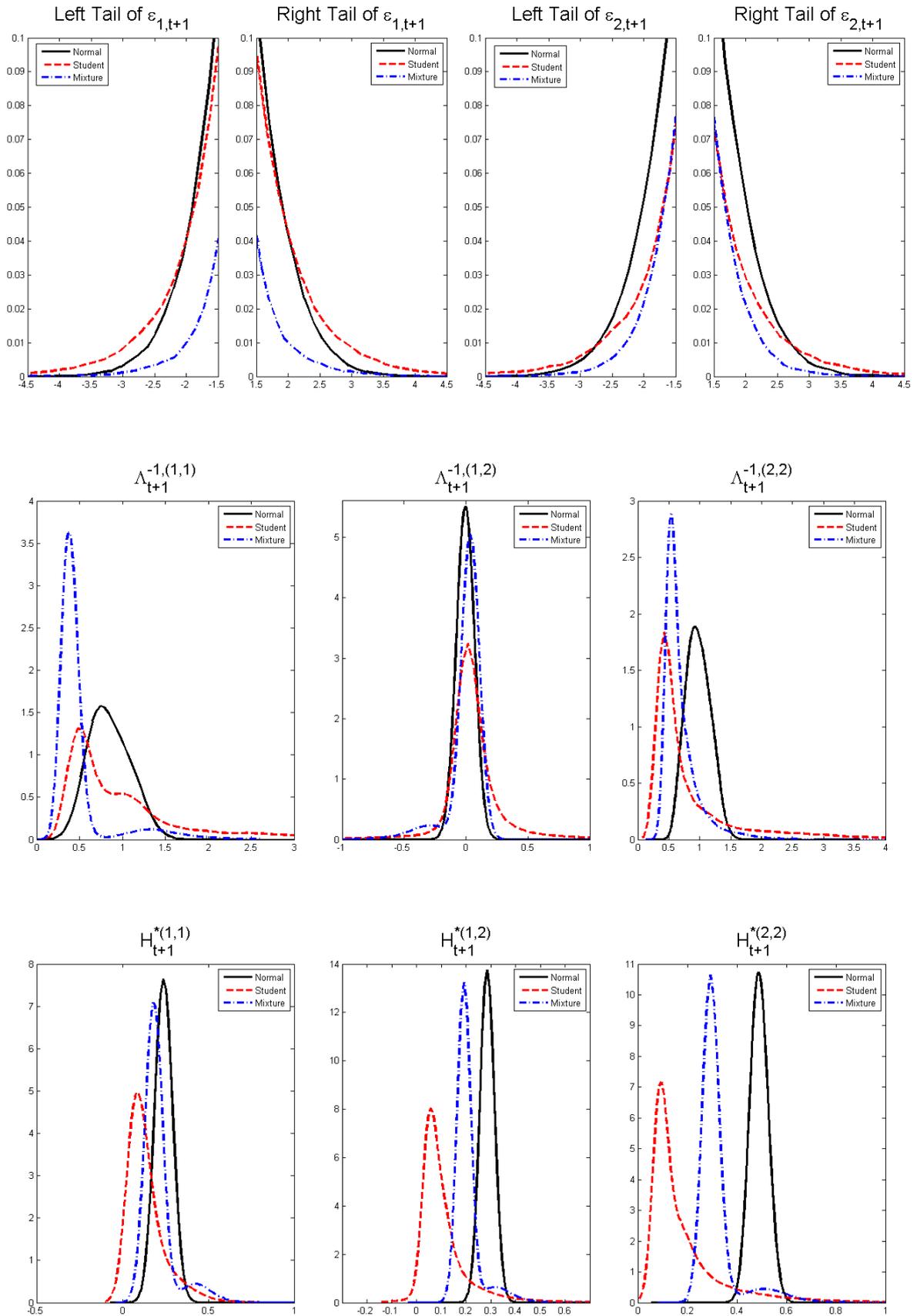
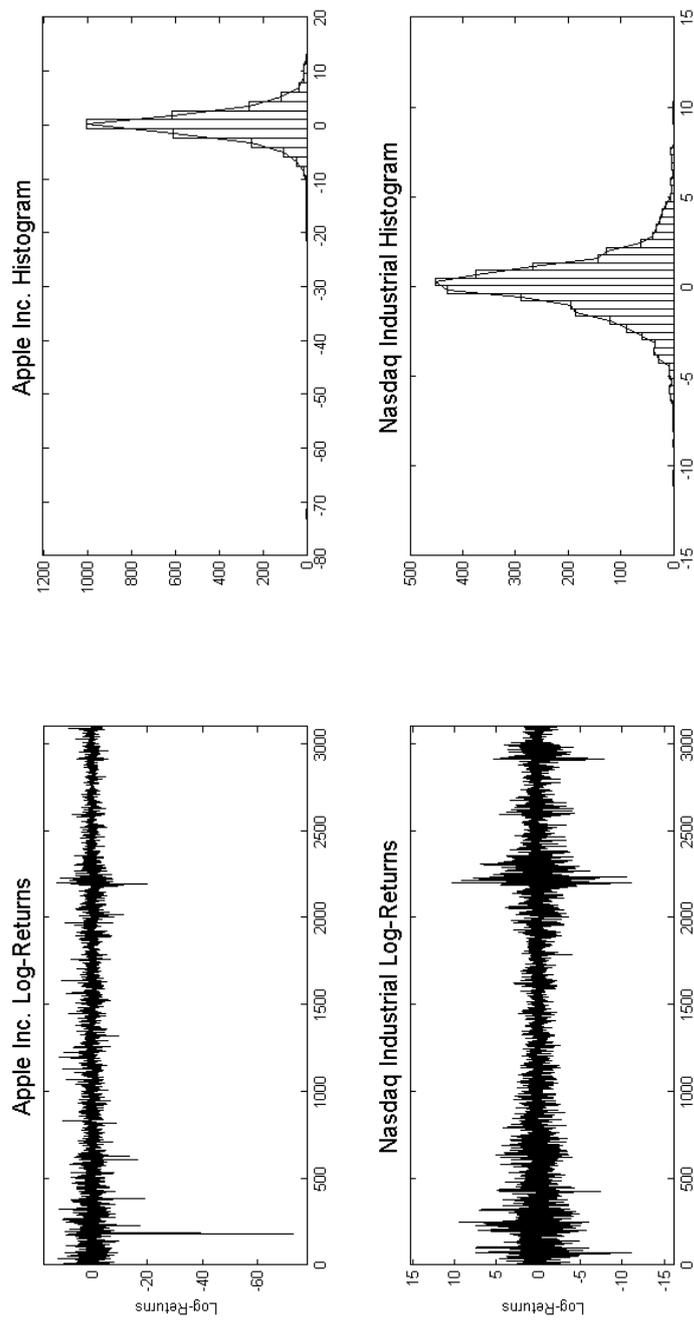
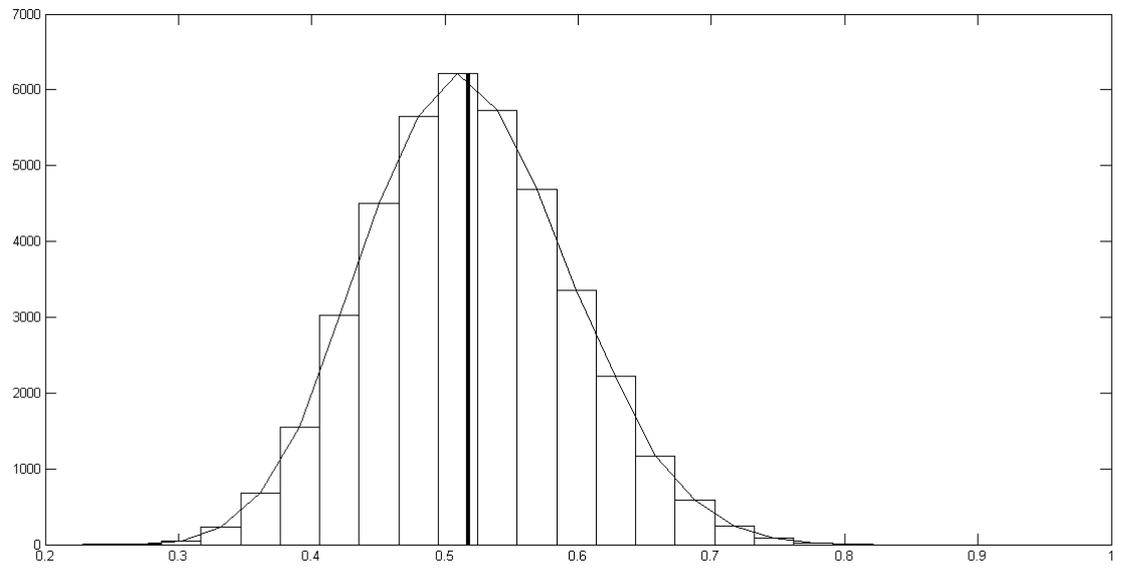
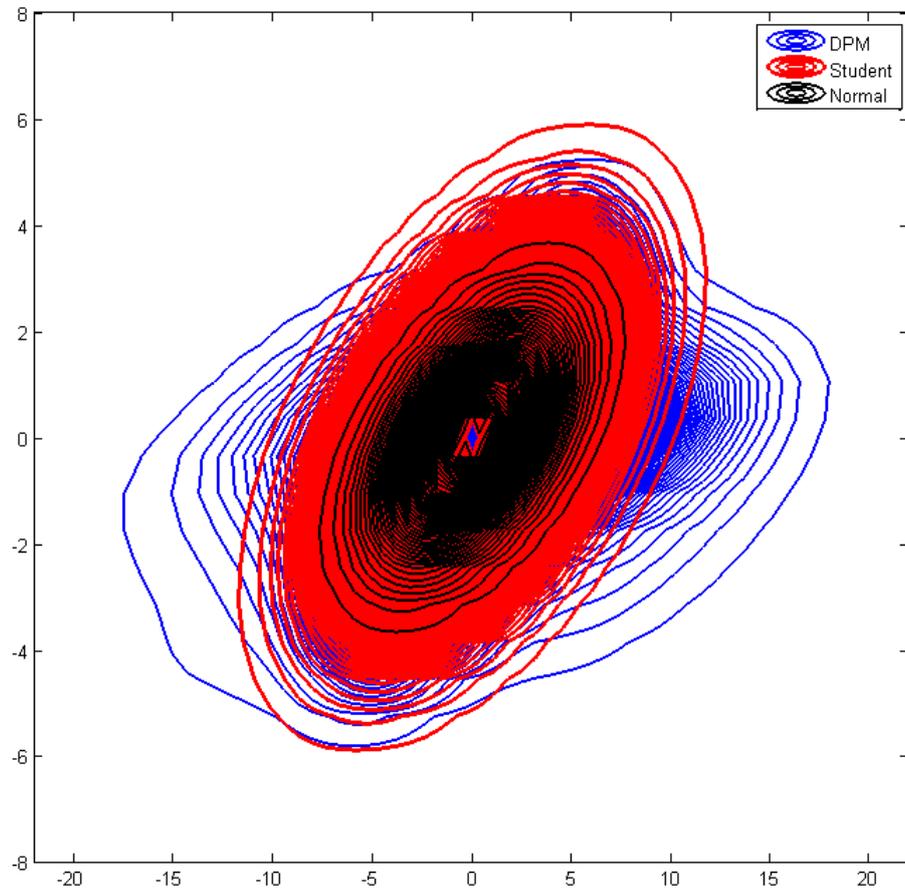
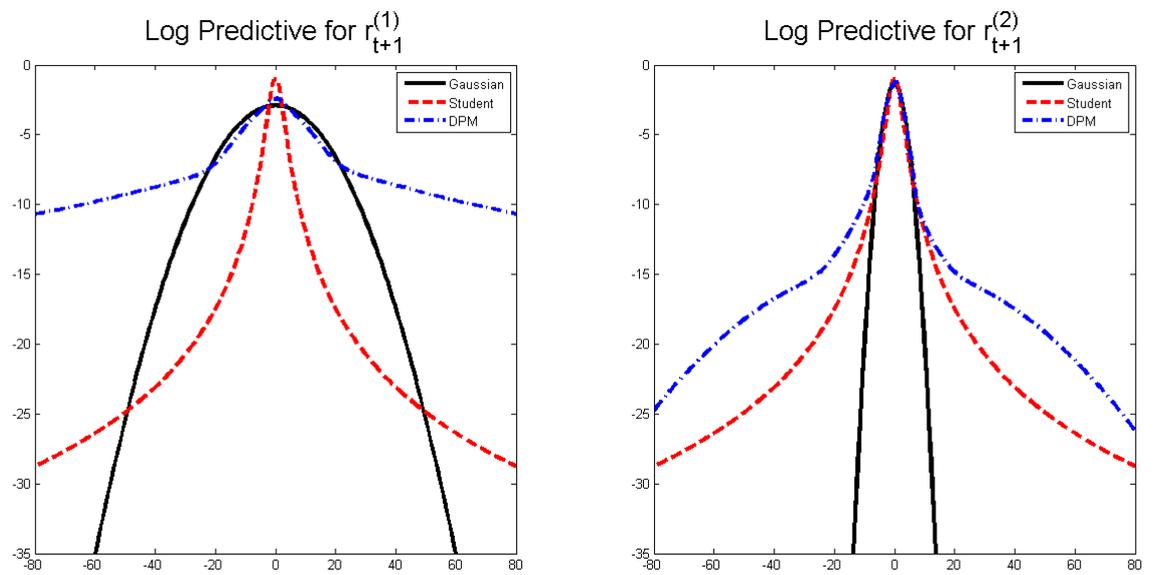


Figure 4. Log&gt;Returns and Histograms of Apple Inc. and NASDAQ Ind. Index

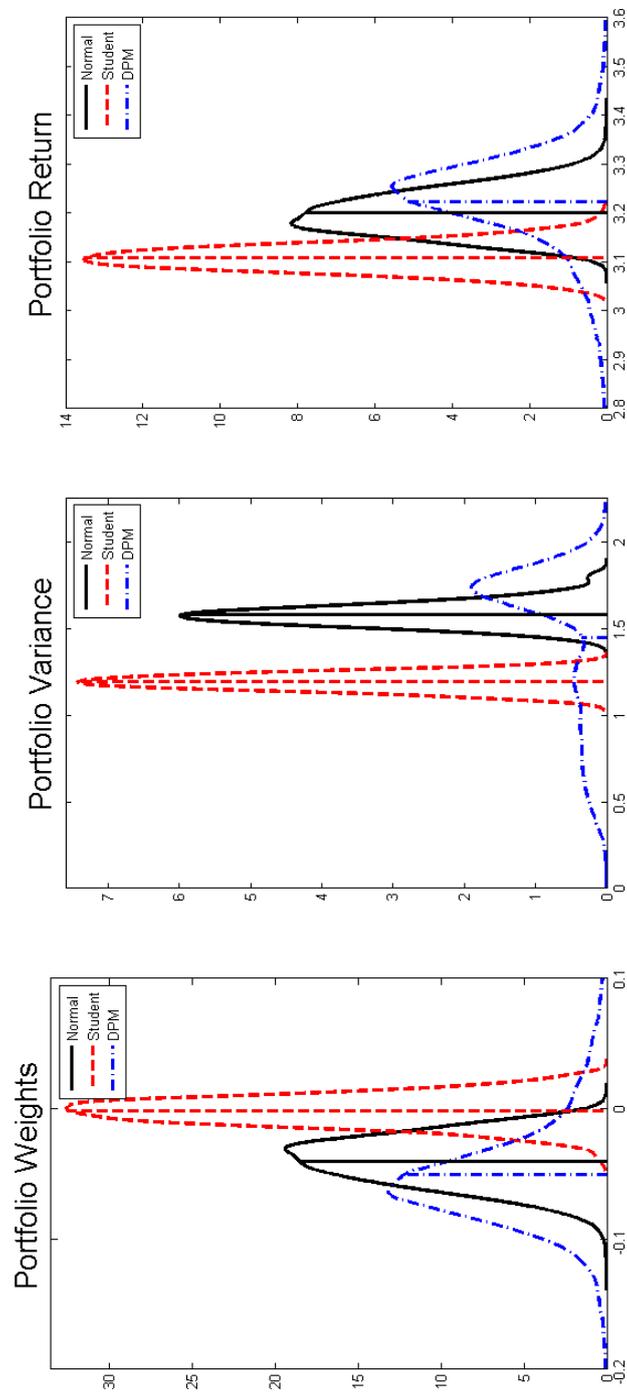


**Figure 5.** Histogram of the  $A = 1/(1 + c)$ 

**Figure 6.** Predictive Contours for  $r_{t+1}$ 

**Figure 7.** Marginal Log Predictive Densities for  $r_{t+1}$ 

**Figure 8.** Kernel Smoothing Densities for  $t + 1$  Portfolio Weights, Variances and Returns



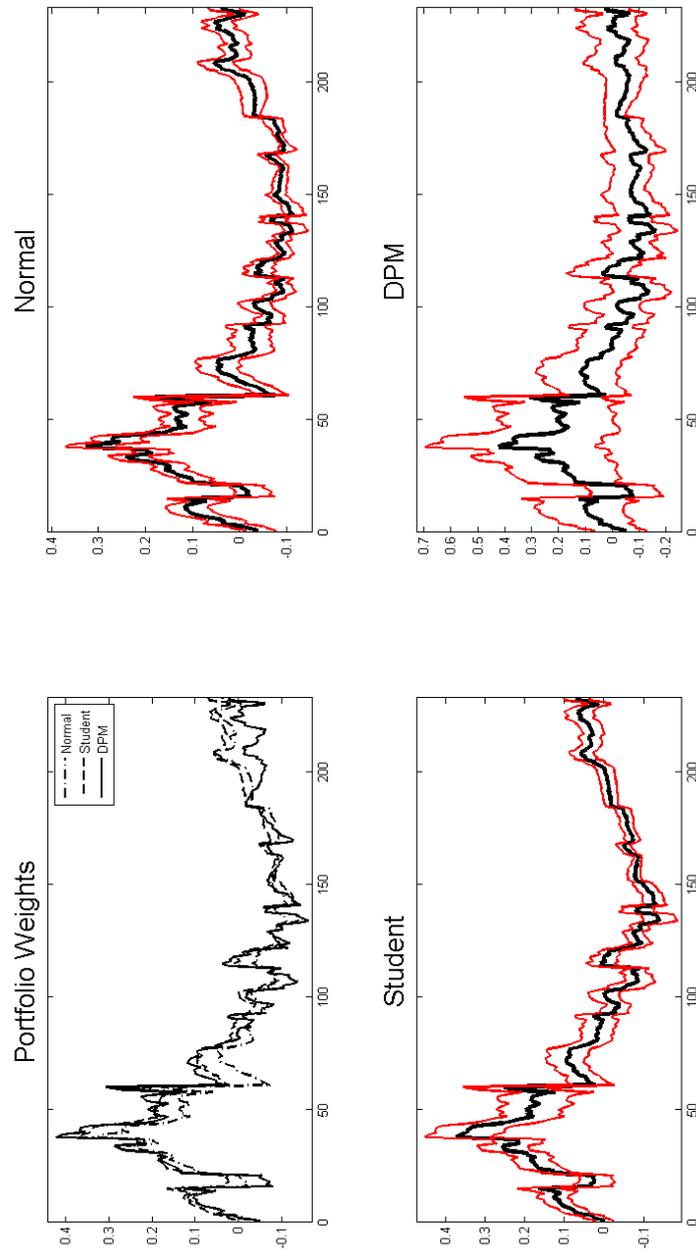
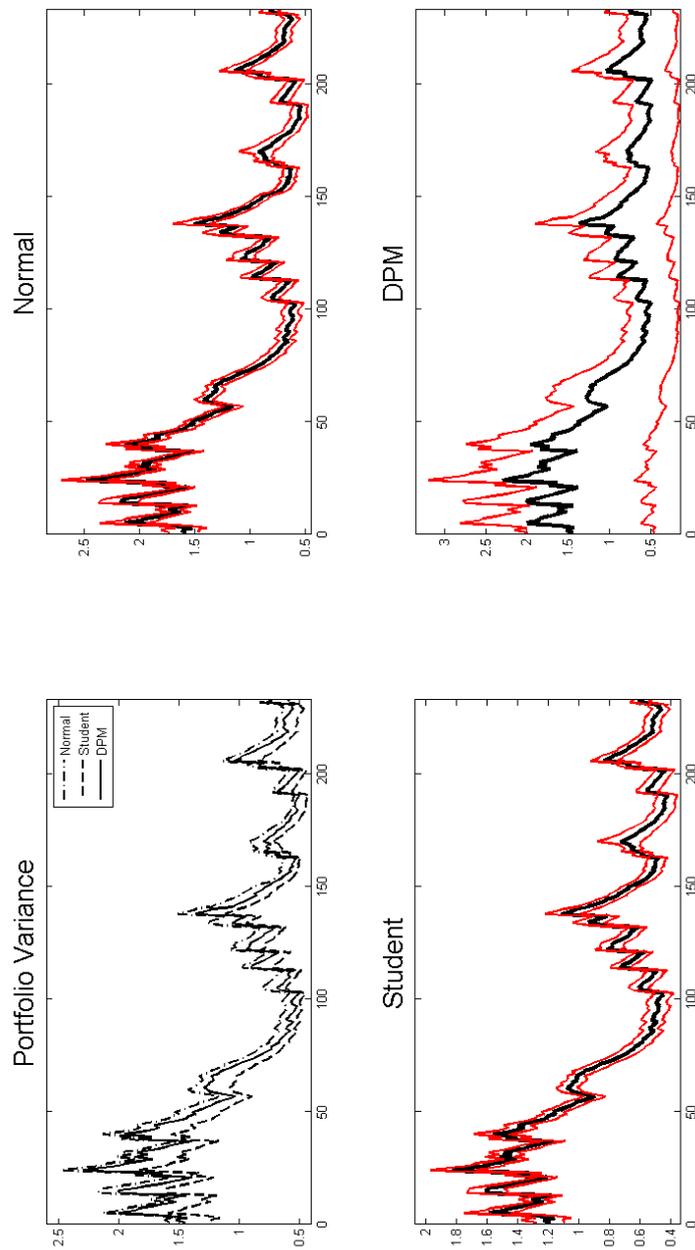
**Figure 9.** Dynamics of Portfolio Weights and 95% Credible Intervals

Figure 10. Dynamics of Optimal Portfolio Variance and 95% Credible Intervals



**Figure 11.** Dynamics of Optimal Portfolio Cumulative Mean Return and 95% Credible Intervals

