

University of Toronto  
Department of Economics



Working Paper 458

Bayesian semiparametric multivariate GARCH modeling

By Mark J Jensen and John M Maheu

June 29, 2012

# Bayesian semiparametric multivariate GARCH modeling\*

Mark J. Jensen<sup>†</sup> and John M. Maheu<sup>‡</sup>

May 2012

## Abstract

This paper proposes a Bayesian nonparametric modeling approach for the return distribution in multivariate GARCH models. In contrast to the parametric literature the return distribution can display general forms of asymmetry and thick tails. An infinite mixture of multivariate normals is given a flexible Dirichlet process prior. The GARCH functional form enters into each of the components of this mixture. We discuss conjugate methods that allow for scale mixtures and nonconjugate methods which provide mixing over both the location and scale of the normal components. MCMC methods are introduced for posterior simulation and computation of the predictive density. Bayes factors and density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach.

---

\*We are grateful to Anatoliy Belaygorod, Sid Chib, James MacKinnon, Bill McCausland, and Benoit Perron for helpful comments and suggestions, and for comments from both the conference participants of the European Seminar on Bayesian Econometrics 2011, CFE'11, the Seminar on Bayesian Inference in Econometrics and Statistics 2012 and the Symposium on Nonlinear Dynamics and Econometrics 2011 and the seminar participants at the University of Montreal and Queen's University. The views expressed here are ours and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Maheu is grateful to the SSHRC for financial support.

<sup>†</sup>Federal Reserve Bank of Atlanta, Mark.Jensen@atl.frb.org

<sup>‡</sup>Department of Economics, University of Toronto, Canada and RCEA, Italy, jmaheu@chass.utoronto.ca.

# 1 Introduction

Modeling comovements in asset returns has been a central topic in asset allocation, risk management, and asset pricing for several decades (for example, see Bollerslev et al. (1988), Buraschi et al. (2010) and Bali & Engle (2010)). Multivariate GARCH (MGARCH) models continue to be the main class of models used in these empirical finance application. Recent surveys of the existing MGARCH models literature can be found in Laurent et al. (2006) and Silvennoinen & Teräsvirta (2009).

Many MGARCH models assume a multivariate normal distribution for the return innovation vector. However, the tails of the empirical distribution of returns are fatter than a normal distribution and less symmetrical. These and other empirical regularities have lead to the rejection of the normality assumption for multivariate returns (see Richardson & Smith (1993)). In response, Diamantopoulos & Vrontos (2010), Harvey et al. (1992) and Fiorentini et al. (2003), adopt a multivariate Student-t density for the MGARCH model, whereas Bauwens & Laurent (2005) extend this to a multivariate skew-Student distribution. In addition, a finite mixture of normal distributions has been applied by Bauwens et al. (2007) and Galeano & Ausín (2010). Each of these extensions improves the fit of the MGARCH but all remain essentially parametric approaches.

Other notable advances are the classical semiparametric approaches considered by Long et al. (2011) and Hafner & Rombouts (2007). Both of these semiparametric MGARCH models employ a two-step estimation strategy. It is unclear what the finite sample performance is for estimators of these more general MGARCH models, or the effect of two-step estimation. On the other hand, an open issue is the asymptotic normality of maximum likelihood estimators (see point 9 of Bauwens et al. (2007)).

An alternative likelihood based approach is the posterior analysis of the Bayesian paradigm. Bayesian methods are attractive in that estimation is conceptually straightforward, but they avoid sample size issues and two-step estimation questions. Existing Bayesian estimators of the MGARCH model by Dellaportas & Vrontos (2007), Hudson & Gerlach (2008) and Osiewalski & Pipien (2004) are based on parametric models. The purpose of this paper is to extend this work to a Bayesian semiparametric setting.

This paper provides Markov chain Monte Carlo (MCMC) posterior simulation methods for Bayesian semiparametric MGARCH models. Estimation is jointly conducted on all parameters of the conditional covariances and the unknown innovation distribution. The predictive density integrates out uncertainty from the conditional covariance as well as the uncertainty about the return distribution.

Our semiparametric MGARCH model is closely related to the semiparametric univariate volatility models of Jensen & Maheu (2010, 2012) and Ausín et al. (2010). These earlier semiparametric models were based on modeling the dynamics of conditional volatility parametrically and on a nonparametric Dirichlet process mixture (DPM) prior for the distribution of the standardized returns. We too assume a parametric model for the dynamics of the conditional variance matrix of returns<sup>1</sup> and a DPM prior for the multivariate distribution of the standardized returns.

The Bayesian nonparametric DPM prior, while being a flexible, countably, infinite mixture of multivariate normals with unknown means, covariances and mixture probabilities, is also parsimonious and simple to sample from. Our first nonparametric approach makes use of the Dirichlet processes Polya urn sampler of Escobar & West (1995) and West et al. (1994) to produce draws of the unknown distribution by sampling the mixture covariance matrices. Posterior draws consists of sampling each components covariance matrix from its conjugate, conditional, posterior distribution.

It is well understood that as the number of assets in the MGARCH model increases, the number of unknowns can become overwhelming. To ensure that our semiparametric Bayesian approach is scalable, we also estimate our semiparametric MGARCH model with a DPM prior restricted to covariance matrices having only diagonal elements. These diagonal covariance matrices take on two forms; one prior having only a single scaling factor and the other allowing each asset to have its own mixture variance.

In order to allow more flexible priors and enable different location vectors in the mixture components, which will be important in capturing potential asymmetries in the nonparametric distribution of the standardized errors, we also use the slice sampler of Walker (2007). Unlike the above Gibbs based, Polya urn approach, the slice sampler works directly with the stick breaking representation of the DP prior. Posterior sampling methods are straightforward and allow inference on the unknown distribution itself. Such inference is not possible with the Gibbs based approach since it generates draws by marginalizing over the DP prior.

Empirical applications to foreign exchange returns and equity returns using a vector-diagonal MGARCH model of Ding & Engle (2001) shows the usefulness of our approach. The semiparametric model has very similar parameter estimates to a MGARCH model with Student-t innovations. Time-series plots of the conditional variances and condi-

---

<sup>1</sup>We choose to apply our Bayesian semiparametric estimator to Ding & Engle (2001) vector-diagonal version of Bollerslev et al. (1988) diagonal VEC GARCH model, but other parametric models of the conditional variance matrix could be used.

tional correlations are likewise similar. We show how the Bayes factor for the parametric Student-t MGARCH model versus the simplest nonparametric specification can be computed using the Savage-Dickey density ratio. For each dataset we find very strong support for the nonparametric mixture model.

To assess our more general model we turn to predictive Bayes factors. We show how to compute the predictive likelihood for the full support of the distribution as well as regions such as the predictive likelihood for the tail of the distribution. In these comparisons the semiparametric model is at least as good as the parametric alternative and often significantly better. For example, in the first equity application the predictive log-Bayes factor in favor of the new model is 17. On average 10 components are used in the mixture to approximate the distribution of returns.

We present period by period log-predictive likelihoods for model comparison. We find that for foreign exchange data the new models essentially match the predictive performance of a MGARCH model with Student-t innovations. For equity the new models offer significant improvements. The gains from the semiparametric model often occur during high volatility episodes. We conclude that our model can be substituted for parametric models and be expected to deliver similar and potentially much better predictive performance.

During the high volatility of the financial crisis the sequential estimates of the semiparametric scale mixture model go from about 9 components in the mixture to around 10. Towards the end of the sample period the number of components begin to decrease. This is a flexible feature of the DPM mixture model. If the future is unlike the past the model can introduce new components with new parameters into the mixture to accommodate new structure in the distribution. This is something that is not possible for the t-distribution or a finite mixture model as the number of components is fixed.

The paper concludes with an application to a larger equity portfolio of 10 assets. Computation becomes more challenging with this model so we apply the diagonal version of the DPM covariance prior to reduce the computational costs. The MGARCH model with a DPM prior mixing over each assets variances leads to better out-of-sample forecast performance. This semiparametric MGARCH model uses approximately 25 mixture components and continues to have significant gains in density forecasts of returns beyond the MGARCH model with Student-t innovations.

This paper is organized as follows. The next section presents a MGARCH model with Student-t innovations that provides a benchmark to compare to the new models. Section 3 removes the Student-t distributional assumption and replaces it with a Dirichlet

process mixture model. Gibbs sampling is discussed and computation of the predictive density for returns and for a portfolio of assets is derived. Section 4 introduces posterior sampling for a DPM model with more general priors for the base measure. We use this version to allow the mean of the components in the infinite mixture to be non-zero. Data, estimation results and model comparison are found in Section 5 while the paper concludes in Section 6.

## 2 Benchmark Model

In the following let  $y_t = (y_{1t}, \dots, y_{kt})'$  denote a vector of  $k$  asset returns. For comparison purposes we consider a parametric version of the vector-diagonal multivariate GARCH model of Ding & Engle (2001). To capture the fat tails so prominent in asset returns a Student-t distribution is used for the conditional density. The model, MGARCH-t, is

$$y_t|H_t \sim St_k(0, H_t, \psi) \quad (1)$$

$$H_t = \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}' + \Gamma_2 \odot H_{t-1}, \quad (2)$$

where  $St_k(0, H_t, \psi)$  is a  $k$  variate multivariate Student-t distribution with mean vector 0,  $k \times k$  scale matrix  $H_t$  and  $\psi$  is the degree of freedom parameter. The conditional second moment is  $\text{Cov}(y_t|H_t, \psi) = \frac{\psi}{\psi-2}H_t$ . Note that  $\Gamma_0$  is a symmetric positive definite matrix parametrized as  $\Gamma_0 = \Gamma_0^{1/2}(\Gamma_0^{1/2})'$  where  $\Gamma_0^{1/2}$  is a lower triangular matrix and  $\Gamma_1 = \gamma_1(\gamma_1)'$  and  $\Gamma_2 = \gamma_2(\gamma_2)'$ , where both  $\gamma_1$  and  $\gamma_2$  are  $k$  vectors. The symbol  $\odot$  denotes the Hadamard product. This model assumes that each conditional second moment element  $h_{ij,t}$  is only related to a lag of itself,  $h_{ij,t-1}$ , plus its past innovation shocks  $y_{i,t-1}y_{j,t-1}$ . The rank one restrictions on  $\Gamma_1$  and  $\Gamma_2$  serve to keep the model parsimonious. To start the GARCH recursion  $H_1$  is set to the sample covariance of  $y_t$ . Several features of this model are discussed in detail in Ledoit et al. (2003).

Given a dataset  $Y_T = (y_1, \dots, y_T)$ , model parameter vector  $\Gamma = \{\Gamma_0^{1/2}, \gamma_1, \gamma_2, \psi\}$  and prior  $p(\Gamma)$  the posterior density is

$$p(\Gamma|Y_T) \propto p(\Gamma) \prod_{t=1}^T \frac{\psi^{\psi/2} \Gamma[(\psi+k)/2]}{\Gamma(\psi/2)} |H_t|^{-1/2} (\psi + y_t' H_t^{-1} y_t)^{-(\psi+k)/2}. \quad (3)$$

The priors are independent for each individual parameter and follow  $N(0, 100)$  with the following restrictions  $\text{Diag}(\Gamma^{1/2}) > 0$ ,  $\gamma_{11} \geq 0$ ,  $\gamma_{22} \geq 0$  to impose identification and  $\psi > 2$ . This density on (3) is nonstandard and we employ a Metropolis-Hastings sampler. Given

the current value  $\Gamma$  of the chain, the proposal  $\Gamma'$  is sampled from

$$h(\Gamma') \sim \begin{cases} N(\Gamma, V) & \text{w.p. } p \\ N(\Gamma, 100V) & \text{w.p. } 1 - p, \end{cases} \quad (4)$$

and accepted with probability

$$\min\{p(\Gamma'|Y_T)/p(\Gamma|Y_T), 1\} \quad (5)$$

and otherwise rejected and  $\Gamma$  is selected as the draw. This is a multivariate random walk directed by a fat-tailed mixture of normals. The second normal in the proposal mixture allows for the possibility of large moves in the parameter space. In the empirical applications  $p = 0.9$  and  $V$  is an estimate of the inverse hessian of  $\log(p(\Gamma|Y_T))$  evaluated at the posterior mode and scaled to achieve an acceptance frequency between 0.2–0.5.  $V$  is computed once at the start of estimation.

Sampling from this distribution is repeated many times and after dropping a burnin sample we collect  $\{\Gamma^{(i)}\}_{i=1}^N$  to estimate posterior quantities of interest. The predictive density can be approximated as

$$p(y_{t+1}|Y_t) \approx \frac{1}{N} \sum_{i=1}^N St_k(y_{t+1}|0, H_{t+1}^{(i)}, \psi^{(i)}) \quad (6)$$

where  $St_k(y_{t+1}|0, H_t^{(i)}, \psi^{(i)})$  denotes the p.d.f. of the Student-t distribution and  $H_{t+1}^{(i)}$  is computed according to (2) with parameter  $\Gamma^{(i)}$ .

### 3 Multivariate GARCH-DPM- $\Lambda$ Model

Now consider the same MGARCH-t model with the Student-t innovations replaced by a Dirichlet process mixture model. This is labelled MGARCH-DPM- $\Lambda$  and follows,

$$y_t|\Lambda_t, H_t \sim N\left(0, H_t^{1/2}\Lambda_t^{-1}(H_t^{1/2})'\right), \quad (7)$$

$$H_t = \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}' + \Gamma_2 \odot H_{t-1} \quad (8)$$

$$\Lambda_t|G \stackrel{iid}{\sim} G, \quad (9)$$

$$G|G_0, \alpha \sim \text{DP}(G_0, \alpha), \quad (10)$$

$$G_0(\Lambda_t) \equiv \text{Wishart}_k(P, v + k - 1), \quad v \geq 1. \quad (11)$$

The MGARCH parameters  $\Gamma_0$ ,  $\Gamma_1$  and  $\Gamma_2$  follow the same structure as the previous sections MGARCH-t model. (9)-(11) places a nonparametric prior on the random innovation distribution of returns. This Bayesian nonparametric prior is an infinite mixture

of multivariate normals with mixing over the covariances and can approximate a wide class of distributions. Within this mixture the unknown matrix  $\Lambda_t$  is assumed to be distributed according to an unknown distribution  $G$ .  $G$  is modeled nonparametrically in terms of the Dirichlet process (DP) prior of Ferguson (1973) with base measure  $G_0(\Lambda_t)$  and precision parameter  $\alpha$ . The DP is centered at  $G_0$  since  $E[G(\Lambda)] = G_0(\Lambda)$  for the measurable set  $\Lambda$ . The parameter  $\alpha$  is a precision parameter in the sense that  $\text{Var}(G(\Lambda)) = G_0(\Lambda)(1 - G_0(\Lambda))/(\alpha + 1)$ . A larger  $\alpha$  is consistent with the belief that  $G$  is closer to  $G_0$ . The base measure is a Wishart density<sup>2</sup> of dimension  $k$  with scale matrix  $P$  and degrees of freedom  $v + k - 1$ . The Wishart is a distribution of symmetric nonnegative-definite random matrices and facilitates Gibbs sampling for the DPM model portion of posterior sampling.

The model can be cast in the Sethuraman (1994) representation of the DPM mixture model as

$$y_t|H_t \sim \sum_{j=1}^{\infty} V_j f_k \left( y_t | 0, H_t^{1/2} D_j^{-1} (H_t^{1/2})' \right) \quad (13)$$

where the mixture only affects the covariance of each component and  $f_k \left( y_t | 0, H_t^{1/2} D_j^{-1} (H_t^{1/2})' \right)$  is a multivariate normal distribution of dimension  $k$  with mean vector 0 and covariance  $H_t^{1/2} D_j^{-1} (H_t^{1/2})'$ . The mixture weights are distributed as  $V_1 = W_1$ , and  $V_j = W_j \prod_{s=1}^{j-1} (1 - W_s)$ , where  $W_j \sim \text{Beta}(1, \alpha)$ . The mixture parameters  $D_j$  prior is found in (11).

The semiparametric model nests two important cases. For example,  $G$  becomes a single point mass as  $\alpha \rightarrow 0$  and  $V_1 = 1, V_2 = 0, V_3 = 0, \dots$ . In this case there is only 1 component in the DPM and we have a Gaussian model for  $y_t$ . On the other hand, as  $\alpha \rightarrow \infty$  we have an infinite number of distinct components all drawn from the Wishart distribution and all equally likely. In this case  $G$  replicates  $G_0$  so that  $G \equiv G_0$ . This gives a Student-t distribution for  $y_t$ .

In this formulation of the DPM prior only the precision matrix  $\Lambda_t$  affects components in the mixture. This allows for Gibbs sampling that is straightforward but further on a non-zero mean vector is allowed.

Given a dataset  $Y_T = \{y_1, \dots, y_T\}$ , the Gibbs sampler will sequentially draw the set

---

<sup>2</sup>If  $A \sim \text{Wishart}_k(S_0, v_0)$  with scale matrix  $S_0$  and degree of freedom  $v_0$  then it has density

$$\text{Wishart}_k(A|S_0, v_0) = 2^{-v_0 k/2} \pi^{-k(k-1)/4} |S_0|^{-v_0/2} \left[ \prod_{i=1}^k \Gamma \left( \frac{v_0 - 1 - i}{2} \right) \right]^{-1} |A|^{(v_0-1-k)/2} \exp \left( -\frac{1}{2} \text{tr}(S_0^{-1} A) \right). \quad (12)$$



$\{\Lambda_1, \dots, \Lambda_T\}$ . Due to the nature of the DPM model some of the  $\Lambda_t$  will be identical. This clustering is one of the attractions of the DPM model as it promotes parsimony. Let  $B_j, j = 1, \dots, m$  denote the  $m < T$  unique clusters of  $\Lambda_t$  and  $B = \{B_1, \dots, B_m\}$ . Observation  $t$  is assigned to cluster  $j$  with parameter  $B_j$  if  $s_t = j$ . Let  $S = (s_1, \dots, s_T)$  and  $S_{-t}$  be the state indicators for  $\Lambda_{-t} = (\Lambda_1, \dots, \Lambda_{t-1}, \Lambda_{t+1}, \dots, \Lambda_T)$  and  $m^{(t)}$  the unique number of clusters for the set  $\Lambda_{-t}$ .

To describe the sampler for  $B, S|Y_T, H_1, \dots, H_T$  we first apply the transformation

$$x_t \equiv H_t^{-1/2} y_t \quad (14)$$

and rewrite Equation (7) so that:

$$x_t|\Lambda_t \sim N(0, \Lambda_t^{-1}) \quad (15)$$

$$\Lambda_t|G \stackrel{iid}{\sim} G, \quad (16)$$

$$G|G_0, \alpha \sim \text{DP}(G_0, \alpha), \quad (17)$$

$$G_0(\Lambda_t) \equiv \text{Wishart}_k(P, v + k - 1), \quad v \geq 1. \quad (18)$$

Standardizing the returns in (14) leaves the return innovation distribution which is modeled as a flexible DPM. Although we focus on one particular MGARCH model other MGARCH parametrizations could be used as long as the transformation in (14) can be applied. The DPM model for  $x_t$  does not in general impose a covariance matrix of  $I_k$ . Imposing moment restrictions in DPM models is still an open question but for our applications we use the prior to essentially center  $x_t$  around a covariance of  $I_k$ .

With this transformation we can use the Gibbs sampling approach of Escobar & West (1995) to sample the DPM model portion of the model. Let  $X_T = \{x_1, \dots, x_T\}$ , then draws are now made from  $B, S|X_T$  with the following two step procedure:

Step 1. Sample  $s_t, t = 1, \dots, T$  and  $m$  from:

$$\begin{aligned} \Lambda_t|x_t, \Lambda_{-t}, S_{-t} &\sim c \frac{\alpha}{\alpha + T - 1} g(x_t) G(d\Lambda_t|x_t) \\ &+ \frac{c}{\alpha + T - 1} \sum_{j=1}^{m^{(t)}} n_j^{(t)} f_k(x_t|0, B_j^{-1}) \delta_{B_j}(\Lambda_t). \end{aligned} \quad (19)$$

where  $g(x_t) \equiv \int f_k(x_t|0, \Lambda_t^{-1} G_0(d\Lambda_t) d\Lambda_t$ .

Step 2. Given the  $S$  and  $m$  from Step 1, sample  $B_j, j = 1, \dots, m$  from:

$$B_j|\{x_t : s_t = j\} \propto \prod_{t:s_t=j} f_k(x_t|0, \Lambda_t^{-1}) G_0(d\Lambda_t). \quad (20)$$

We require the following derivations.

$$G(d\Lambda_t|x_t) \equiv \frac{f_k(x_t|0, \Lambda_t^{-1}) G_0(d\Lambda_t)}{g(x_t)}, \quad (21)$$

$$\propto |\Lambda_t|^{1/2} \exp\left(-\frac{1}{2}x_t' \Lambda_t x_t\right) |\Lambda_t|^{(v+k-1-k)/2} \exp\left(-\frac{1}{2}\text{tr}(P^{-1} \Lambda_t)\right) \quad (22)$$

$$\propto |\Lambda_t|^{(v-1)/2} \exp\left(-\frac{1}{2}\text{tr}([P^{-1} + x_t x_t'] \Lambda_t)\right) \quad (23)$$

$$\propto \text{Wishart}_k(\Lambda_t | (x_t x_t' + P^{-1})^{-1}, v+k), \quad (24)$$

and

$$g(x_t) \equiv \int f_k(x_t|0, \Lambda_t^{-1}) G_0(d\Lambda_t) d\Lambda_t \quad (25)$$

$$\propto |x_t x_t' + P^{-1}|^{-(v+k)/2} \quad (26)$$

$$\propto (v + x_t'(Pv)x_t)^{-(v+k)/2} \quad (27)$$

$$\propto St_k(x_t|0, (Pv)^{-1}, v). \quad (28)$$

In Step 2 we have for  $j = 1, \dots, m$ ,

$$\begin{aligned} B_j|\{x_t : s_t = j\} &\propto \left[ \prod_{t:s_t=j} |B_j|^{1/2} \exp\left(-\frac{1}{2}x_t' B_j x_t\right) \right] |B_j|^{(v-2)/2} \exp\left(-\frac{1}{2}\text{tr}(P^{-1} B_j)\right) \\ &\propto |B_j|^{(v+n_j-2)/2} \exp\left(-\frac{1}{2}\text{tr}\left(\left[\sum_{t:s_t=j} x_t x_t' + P^{-1}\right] B_j\right)\right) \end{aligned} \quad (29)$$

$$\propto \text{Wishart}_k\left(B_j \left| \left(\sum_{t:s_t=j} x_t x_t' + P^{-1}\right)^{-1}, v+k+n_j-1\right.\right), \quad (30)$$

where  $n_j = \#\{t : s_t = j\}$ . This completes the DPM model portion of the Gibbs sampling. Finally, if the precision parameter  $\alpha$  is assigned a Gamma prior Gibbs sampling following Escobar & West (1995) can be used for the conditional posterior distribution.

Next we sample the MGARCH model parameters. Let  $\Gamma = \{\Gamma_0^{1/2}, \gamma_1, \gamma_2\}$ ; note, the semiparametric model does not contain the degree of freedom parameter  $\psi$ . To sample from  $\Gamma|B, S$  note that

$$\text{Cov}(y_t|H_t, B_{s_t}) = H_t^{1/2} B_{s_t}^{-1} (H_t^{1/2})', \quad (31)$$

so that the conditional posterior is

$$p(\Gamma|B, S, Y_T) \propto p(\Gamma) \prod_{t=1}^T |H_t|^{-1/2} |B_{s_t}|^{1/2} \exp\left(-\frac{1}{2}y_t'(H_t^{-1/2})' B_{s_t}^{-1} H_t^{-1/2} y_t\right). \quad (32)$$

The same sampling strategy used for the MGARCH-t model with a fat-tailed mixture of normals (4)-(5) is used to jointly sample all structural parameters in the MGARCH recursion.

Sampling from these distributions is repeated many times and after dropping a burnin sample we collect  $\{\Gamma^{(i)}, B^{(i)}, S^{(i)}\}_{i=1}^N$  to estimate posterior quantities of interest.

### 3.1 Predictive Density

Escobar & West (1995) point out that the key quantity of interest for Bayesian nonparametrics is the predictive density. The predictive density for  $x_{t+1}$  given the parameters follows the Polya urn prediction rule,

$$p(x_{t+1}|X_t, \Gamma, B, S) = \frac{\alpha}{\alpha + t} St_k(x_{t+1}|0, (P\nu)^{-1}, \nu) + \sum_{j=1}^m \frac{n_j}{\alpha + t} f_k(x_{t+1}|0, B_j^{-1}). \quad (33)$$

Applying the change of variables  $y_{t+1} = H_{t+1}^{1/2} x_{t+1}$ , with Jacobian term  $|H_{t+1}|^{-1/2}$ , gives the predictive density for  $y_{t+1}$  as,

$$\begin{aligned} p(y_{t+1}|Y_t, \Gamma, B, S) &= \frac{\alpha}{\alpha + t} St_k\left(y_{t+1}|0, H_{t+1}^{1/2}(P\nu)^{-1}(H_{t+1}^{1/2})', \nu\right) \\ &+ \sum_{j=1}^m \frac{n_j}{\alpha + t} f_k\left(y_{t+1}|0, H_{t+1}^{1/2} B_j^{-1} (H_{t+1}^{1/2})'\right). \end{aligned} \quad (34)$$

From this the covariance is

$$\text{Cov}(y_{t+1}|Y_t, \Gamma, B, S) = \frac{\alpha}{\alpha + t} \frac{H_{t+1}^{1/2} P^{-1} (H_{t+1}^{1/2})'}{\nu - 2} + \sum_{j=1}^m \frac{n_j}{\alpha + t} H_{t+1}^{1/2} B_j^{-1} (H_{t+1}^{1/2})'. \quad (35)$$

The predictive density with all parameter and density uncertainty integrated out can be estimated as

$$p(y_{t+1}|Y_t) \approx \frac{1}{N} \sum_{i=1}^N p(y_{t+1}|Y_t, \Gamma^{(i)}, B^{(i)}, S^{(i)}) \quad (36)$$

where  $\{\Gamma^{(i)}, B^{(i)}, S^{(i)}\}_{i=1}^N$  are the posterior draws and the right-hand side of (36) averaging over (33).

The predictive density of a portfolio of assets can also be conveniently derived from these results. Given the weights  $\omega$  on wealth with  $\sum_{i=1}^k \omega_i = 1$ , the portfolio return is  $y_{t+1}^p = \omega' y_{t+1}$  and has a predictive density

$$\begin{aligned} p(y_{t+1}^p|Y_t, \Gamma, B, S, \omega) &= \frac{\alpha}{\alpha + t} St_1\left(y_{t+1}^p|0, \omega' H_{t+1}^{1/2} (P\nu)^{-1} (H_{t+1}^{1/2})' \omega, \nu\right) \\ &+ \sum_{j=1}^m \frac{n_j}{\alpha + t} f_1\left(y_{t+1}^p|0, \omega' H_{t+1}^{1/2} B_j^{-1} (H_{t+1}^{1/2})' \omega\right). \end{aligned} \quad (37)$$

Each of the distributions in this mixture are univariate. Averaging over the posterior draws as in (36) gives  $p(y_{t+1}^p|Y_t, \omega)$ .

### 3.2 Savage-Dickey Density Ratio

The DP precision parameter  $\alpha$  can be understood as the parameter controlling the number of unique  $\Lambda_t$ s. Its value determines the mixture weights through the proportion  $W_j \sim \text{Beta}(1, \alpha)$  with expected value  $E[W_j] = 1/(1 + \alpha)$ . As  $\alpha \rightarrow \infty$ , the DP mixture probabilities  $V_j \rightarrow 0$ ,  $G|G_0, \alpha \equiv G_0$ , and  $\Lambda_t \sim G_0$ . Under this condition the MGARCH-DPM- $\Lambda$  is the MGARCH-t model.

As a sharp hypothesis on  $\alpha$ , the Savage-Dickey density ratio of Dickey (1971) can be used to compute the Bayes factor for MGARCH-t versus MGARCH-DPM- $\Lambda$  model. Under a proper prior for  $\alpha$ , the Bayes factor can be computed using the Savage-Dickey density ratio

$$BF \equiv \frac{p(Y_T|\text{MGARCH-t})}{p(Y_T|\text{MGARCH-DPM-}\Lambda)} \quad (38)$$

$$= \frac{p(\alpha \rightarrow \infty|Y_T)}{p(\alpha \rightarrow \infty)} \quad (39)$$

where  $p(Y_T|M)$  is the marginal likelihood of the model  $M$ .

Since  $\alpha$  does not lend itself to a sharp hypothesis we transform the DP precision variable into

$$u \equiv \frac{\alpha}{\alpha + 1}$$

and assume a uniform prior over the unit interval for  $u$ . Using this transformation  $u \rightarrow 1$  as  $\alpha \rightarrow \infty$ . It also provides an intuitive look into how the prior for  $\alpha$  is a prior over the probability of clustering since  $u$  is the probability of a new cluster for the second observation. If  $u = 1$  a new  $\Lambda_t$  will be drawn from  $G_0$  for every observation and  $x_t$  in (15) is distributed as a multivariate Student-t,  $St(0, (P\nu)^{-1}, \nu)$ .

Since for the uniform prior  $p(u = 1)$  equals one, the Savage-Dickey density ratio can be rewritten as

$$BF = p(u = 1|Y_T). \quad (40)$$

The uniform prior on  $u$  implies the following prior for  $\alpha$

$$p(\alpha) = \frac{1}{(\alpha + 1)^2}, \quad \alpha \geq 0.$$

A random walk proposal is used to generate the draw  $\alpha'$ . This candidate draw,  $\alpha'$ , is accepted with probability

$$\frac{p(m|\alpha', T)p(\alpha')}{p(m|\alpha, T)p(\alpha)} = \frac{\alpha'^m \Gamma(\alpha') / \Gamma(\alpha' + T) p(\alpha')}{\alpha^m \Gamma(\alpha) / \Gamma(\alpha + T) p(\alpha)} \quad (41)$$

where  $\alpha$  is the draw from the previous sweep and  $m$  is the number of clusters from the current sweep (see Escobar & West (1995) for the formula of the likelihood function,  $p(m|\alpha, N)$ ). For each draw of  $\alpha$  we also compute its corresponding  $u$ . We then evaluate the empirical distribution of the sampled  $u$  at 1.0. If it is less than 1 then evidence in favor of the MGARCH-DPM- $\Lambda$  model over the Student-t version of the model.

## 4 General Semiparametric MGARCH Model

The previous model presented relies on a restrictive prior and does not allow the mean of a component to affect the mixture. To allow for more general priors on the base measure including estimation of the mean of the components, a sampler based on the stick breaking formulation of the DPM model is used. This more general MGARCH model is labelled as MGARCH-DPM.

The MGARCH-DPM model is,

$$y_t | \mu_t, \Lambda_t, H_t \sim N(H_t^{1/2} \mu_t, H_t^{1/2} \Lambda_t^{-1} (H_t^{1/2})'), \quad (42)$$

$$H_t = \Gamma_0 + \Gamma_1 \odot y_{t-1} y_{t-1}' + \Gamma_2 \odot H_{t-1} \quad (43)$$

$$\mu_t, \Lambda_t | G \stackrel{iid}{\sim} G, \quad (44)$$

$$G | G_0, \alpha \sim \text{DP}(G_0, \alpha), \quad (45)$$

$$G_0(\mu_t, \Lambda_t) \equiv N(\underline{\beta}, \underline{V}) - \text{Wishart}_k(P, v + k - 1), \quad v \geq 1. \quad (46)$$

The base measure governing  $\mu_t$  and  $\Lambda_t$  are independent multivariate normal and Wishart distributions. The mixture parameters now affect the mean vector and covariance in (42).

The Polya-urn based sampler relies on  $G$  being integrated out from the posterior. This marginal method gives a finite number of parameters that can be sampled with MCMC methods when conjugacy is available. For more general priors this does not work. Instead we work with conditional sampling approaches that directly sample from the Sethuraman representation of the DPM. The problem is that there is an infinite number of parameters to keep track of. One approach is to approximate the infinite expansion by a finite one as in Ishwaran & Zarepour (2000).

However, it is possible to avoid this approximation by using slice sampling. The key idea behind slice sampling is to introduce a latent variable that slices the infinite expansion in the Sethuraman representation to a finite number of parameters that can be sampled by MCMC. Integrating over the latent variable yields the correct mixture. We extend the slice sampling techniques of Walker (2007), Kalli et al. (2011) and Paspiliopoulos (2008) to our MGARCH-DPM model.

Let  $G \sim DP(G_0, \alpha)$ , whose stick breaking representation is

$$G = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}, \quad (47)$$

where  $\theta_j \stackrel{iid}{\sim} G_0$ ,  $\theta_j = (\xi_j, B_j)$  and  $\xi_j$  and  $B_j$  are the unique mass points of the vector  $\mu_t$  and matrix  $\Lambda_t$ . The weights are

$$w_1 = v_1, \quad w_j = v_j \prod_{i=1}^{j-1} (1 - v_i) \quad (48)$$

with  $v_j \stackrel{iid}{\sim} \text{Beta}(1, \alpha)$ .

Using the transformation  $x_t \equiv H_t^{-1/2} y_t$  again, the density for  $x_t$  is

$$f(x_t | W, \Theta) = \sum_{j=1}^{\infty} w_j f_k(x_t | \xi_j, B_j^{-1}). \quad (49)$$

where  $W = (w_1, w_2, \dots)$  and  $\Theta = \{\theta_1, \theta_2, \dots\}$ . Introducing the auxiliary latent variable  $u_t$ , we define the joint conditional density of  $x_t$  and  $u_t$  as,

$$f(x_t, u_t | W, \Theta) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < w_j) f_k(x_t | \xi_j, B_j^{-1}). \quad (50)$$

The conditional distribution of components that have a weight  $w_j$  less than the slice variable  $u_t$  is 0. If  $u_t$  has a uniform distribution then integration of  $f(x_t, u_t | W, \Theta)$  with respect to  $u_t$  gives  $f(x_t | W, \Theta)$ . On the other hand the marginal density of  $u_t$  is  $\sum_{j=1}^{\infty} \mathbf{1}(u_t < w_j)$ .

Let  $s_t = j$  assign observation  $x_t$  to data density  $f_k(x_t | \xi_j, B_j^{-1})$  with parameter  $\theta_j = (\xi_j, B_j)$ , and  $U = (u_1, \dots, u_T)$ , then the full likelihood is

$$p(X_T, U, S | W, \Theta) = \prod_{t=1}^T f(x_t, u_t, s_t | W, \Theta) = \prod_{t=1}^T \mathbf{1}(u_t < w_{s_t}) f_k(x_t | \xi_{s_t}, B_{s_t}^{-1}) \quad (51)$$

and the joint posterior is

$$p(W_K) \left[ \prod_{i=1}^K p(\xi_i, B_i) \right] \prod_{t=1}^T \mathbf{1}(u_t < w_{s_t}) f_k(x_t | \xi_{s_t}, B_{s_t}^{-1}), \quad (52)$$

where  $W_K = (w_1, \dots, w_K)$  and  $K$  is the smallest natural number that satisfies the condition  $\sum_{j=1}^K w_j > 1 - \min\{u_t\}$ .<sup>3</sup>

The Gibbs sampler samples the following set of variables each iteration,

$$\{(\xi_j, B_j, v_j), j = 1, 2, \dots, K; (s_t, u_t), t = 1, \dots, T\}. \quad (53)$$

After dropping the burnin phase we collect  $i = 1, \dots, N$  samples.

The sampling steps are:

1.  $\pi(\xi_j, B_j | X_T, S) \propto p(\xi_j, B_j) \prod_{\{t: s_t=j\}} f_k(x_t | \xi_j, B_j^{-1})$ ,  $j = 1, \dots, K$ .
2.  $\pi(v_j | S) \propto \text{Beta}(v_j | a_j, b_j)$ ,  $j = 1, \dots, K$ , where

$$a_j = 1 + \sum_{t=1}^T \mathbf{1}(s_t = j), \quad b_j = \alpha + \sum_{t=1}^T \mathbf{1}(s_t > j). \quad (54)$$

3.  $\pi(u_t | W_K, S) \propto \mathbf{1}(0 < u_t < w_{s_t})$ ,  $t = 1, \dots, T$ .
4. Find the smallest  $K$  such that  $\sum_{j=1}^K w_j > 1 - \min\{u_t\}$ .
5.  $P(s_t = j | X_T, \theta, U, W_K) \propto \sum_{j=1}^K \mathbf{1}(w_j > u_t) f_k(x_t | \xi_j, B_j^{-1})$ .

In step 4 additional  $w_j$  and  $\theta_j$  will need to be generated from the prior if  $K$  is incremented.

From step 1 the conditional posterior for  $\xi = \{\xi_1, \dots, \xi_K\}$  is

$$p(\xi | B, S) \propto \prod_{j=1}^K \left[ \prod_{t: s_t=j} f_k(x_t | \xi_j, B_j^{-1}) \right] p(\xi_j). \quad (55)$$

For each  $j$  in which at least one observation is assigned we have

$$\xi_j | X_T, B, S \sim N(\bar{\beta}, \bar{V}) \quad (56)$$

$$\bar{\beta} = \bar{V} \left( \underline{V}^{-1} \underline{\beta} + B_j \sum_{t: s_t=j} x_t \right) \quad (57)$$

$$\bar{V} = (\underline{V}^{-1} + n_j B_j)^{-1}. \quad (58)$$

---

<sup>3</sup>Hjort et al. (2010) list another condition  $\prod_{j=1}^K (1 - v_j) < \min\{u_t\}$ .

For the conditional posterior of  $B = \{B_1, \dots, B_K\}$  we have

$$p(B|X_T, \xi, S) \propto \prod_{j=1}^K \left[ \prod_{t:s_t=j} f_k(x_t|\xi_j, B_j^{-1}) \right] p(B_j). \quad (59)$$

Conjugacy makes this a series of Gibbs draws,

$$B_j | \{x_t : s_t = j\} \sim \text{Wishart}_k \left( \left( \sum_{t:s_t=j} (x_t - \xi_j)(x_t - \xi_j)' + P^{-1} \right)^{-1}, v + k + n_j - 1 \right).$$

For the empty sets  $\{t : s_t = j\}$  in which none of the observations are allocated to the cluster a direct draw from the Normal-Wishart (46) prior for  $\xi_j$  and  $B_j$  is taken.

## 4.1 Predictive density

Conditional on  $H_{t+1}$  the predictive density can be estimated as

$$p(y_{t+1}|Y_t, H_{t+1}) \approx \frac{1}{R} \sum_{i=1}^R f_k \left( y_{t+1} \left| H_{t+1}^{1/2} \xi^{(i)}, H_{t+1}^{1/2} (B^{(i)})^{-1} (H_{t+1}^{1/2})' \right. \right). \quad (60)$$

We found  $R = 2$  produces accurate results. The parameter  $\theta^{(i)} = (\xi^{(i)}, B^{(i)})$  is found from these steps:

1.  $a \sim U(0, 1)$
2. If  $\sum_{j=0}^{\ell-1} w_j^{(i)} < a < \sum_{j=0}^{\ell} w_j^{(i)}$ , with  $w_0 = 0$  and  $\ell \leq K$  then set  $\theta^{(i)} = \theta_\ell^{(i)}$ .
3. If  $a \geq \sum_{j=0}^{K(i)} w_j^{(i)}$  then generate  $\theta^{(i)} \sim G_0$ .

From this the full predictive density with all parameter uncertainty integrated out is obtained as

$$p(y_{t+1}|Y_t) \approx \frac{1}{N} \sum_{i=1}^N p \left( y_{t+1} \left| Y_t, H_{t+1}^{(i)} \right. \right) \quad (61)$$

where  $H_{t+1}^{(i)}$  is obtained from the posterior draw  $\Gamma^{(i)}$  using the MGARCH recursion (43).

As in Section 3.1 the predictive density of the portfolio return  $y_{t+1}^p = \omega' y_{t+1}$  can be computed. The predictive density analogous to (60) is

$$p(y_{t+1}^p|Y_t, H_{t+1}, \omega) \approx \frac{1}{R} \sum_{i=1}^R f_k \left( y_{t+1} \left| \omega' H_{t+1}^{1/2} \xi^{(i)}, \omega' H_{t+1}^{1/2} (B^{(i)})^{-1} (H_{t+1}^{1/2})' \omega \right. \right) \quad (62)$$



where  $\theta^{(i)} = (\xi^{(i)}, B^{(i)})$  comes from steps 1-3 above. The full predictive density with all parameter uncertainty integrated out is

$$p(y_{t+1}^p | Y_t, \omega) \approx \frac{1}{N} \sum_{i=1}^N p\left(y_{t+1}^p \middle| Y_t, H_{t+1}^{(i)}, \omega\right). \quad (63)$$

It is often of interest to focus on density forecasts of a particular region of the distribution. The predictive density for the tail (Diks et al. 2011) of the portfolio is defined for  $y_{t+1}^p < \eta$ ,  $\eta \in \mathbb{R}$  as

$$p(y_{t+1}^p | y_{t+1}^p < \eta, Y_t) = \frac{p(y_{t+1}^p | Y_t) \mathbf{1}_{y_{t+1}^p < \eta}}{\int_{-\infty}^{\eta} p(v_{t+1}^p | Y_t) dv_{t+1}^p} \quad (64)$$

$$\approx \frac{\frac{1}{NR} \sum_{j=1}^N \sum_{i=1}^R f_1(y_{t+1}^p | \omega' H_{t+1}^{(j)1/2} \mu^{(i)}, \omega' H_{t+1}^{(j)1/2} B^{(i)-1} (H_{t+1}^{(j)1/2})' \omega) \mathbf{1}_{y_{t+1}^p < \eta}}{\frac{1}{NR} \sum_{j=1}^N \sum_{i=1}^R \int_{-\infty}^{\eta} f_1(v_{t+1}^p | \omega' H_{t+1}^{(j)1/2} \mu^{(i)}, \omega' H_{t+1}^{(j)1/2} B^{(i)-1} (H_{t+1}^{(j)1/2})' \omega) dv_{t+1}^p} \quad (65)$$

$$= \frac{\frac{1}{NR} \sum_{j=1}^N \sum_{i=1}^R f_1(y_{t+1}^p | \omega' H_{t+1}^{(j)1/2} \mu^{(i)}, \omega' H_{t+1}^{(j)1/2} B^{(i)-1} (H_{t+1}^{(j)1/2})' \omega) \mathbf{1}_{y_{t+1}^p < \eta}}{\frac{1}{NR} \sum_{j=1}^N \sum_{i=1}^R \Phi\left((\eta - \omega' H_{t+1}^{(j)1/2} \mu^{(i)}) / \sqrt{\omega' H_{t+1}^{(j)1/2} B^{(i)-1} (H_{t+1}^{(j)1/2})' \omega}\right)} \quad (66)$$

The denominator is an integration constant that ensures the conditional predictive density integrates to one while  $\Phi(\cdot)$  denotes the cumulative distribution function for the univariate standard normal and  $f_1(\cdot)$  is the associated pdf. In contrast to the previous predictive likelihood that assessed the quality of the predictive density for the whole support of returns, this focuses only on portfolio values less than  $\eta$ . Analogous results apply for the MGARCH-t and the MGARCH-DPM- $\Lambda$  models.

## 5 Application

Two datasets, one of equity and the other foreign exchange are used to estimate the models. Data for equity returns is on IBM, the Center of Research in Security Prices (CRSP) value-weighted market portfolio and HP, all obtained from CRSP for 2001/01/02 – 2009/12/31 (2263 observations). Data from the FX market is for log-returns on Euro-USD, UK-USD, JPY-USD, 1999/01/05 – 2010/04/19 (2834 observations). Table 1 displays summary statistics for daily returns for the datasets along with the sample correlations. Each of the assets display skewness and excess kurtosis. The time series of equity returns is reported in Figure 1 with clear evidence from 2008 on of an increase in volatility from the financial crisis.

For priors, each element of  $\Gamma_0^{1/2}$ ,  $\gamma_1$  and  $\gamma_2$  is independent  $N(0, 100)$  with the first element of each matrix (vector) restricted to be positive to ensure identification. In

the MGARCH-t model  $\nu \sim U(2, 100)$ . For the DPM models,  $\alpha \sim \text{Gamma}(2, 8)$ ,  $P = \frac{1}{\nu+k-1}I_k$  and  $\nu = 10$ . This implies  $E[\Lambda_t] = I_k$  and essentially centers the DPM model around a Gaussian distribution with mean vector 0 and covariance matrix  $I$ .

## 5.1 Estimates

In posterior simulation, for each model, a total of 13000 draws are collected and the first 3000 are dropped as burnin with the remainder being used to estimate posterior features. Posterior draws are shown in Figure 2 for the the MGARCH-DPM- $\Lambda$  model. Although there is some autocorrelation the chain overall mixes well and fully explores the posterior density. In the following we focus on the results for the equity dataset.

Table 2 displays the posterior mean and 0.95 density intervals for several models for the equity data. The first model is a MGARCH with Student-t innovations. The second model MGARCH-DPM- $\Lambda$ , is the MGARCH with a DPM model return distribution which allow the covariance of the components to differ. The last model MGARCH-DPM allows mixing over both mean and covariance of each mixture component.

The structural parameters,  $\text{vech}(\Gamma_0^{1/2})$ ,  $\gamma_1$ , and  $\gamma_2$ , of the MGARCH models, are all very similar across the semiparametric and parametric models. This is especially true for the MGARCH model's estimates of  $\gamma_2$ . Clearly, volatility clustering remains very important in the semiparametric models. In general the parameters are less precise in the semiparametric models with wider density intervals.

The parametric MGARCH-t model features thick tails with the degree of freedom parameter,  $\psi$ , estimated to be 7.7. A drawback of the Student-t density is that a single degree of freedom parameter governs tail thickness in all directions of the density. This is not the case for the semiparametric alternatives. The semiparametric models capture any deviations from the normal distribution by using approximately 9 and 10 components, on average, in the DPM models. However, there is some posterior uncertainty as to the number of components,  $m$ , with a density interval of (5, 16) for the MGARCH-DPM- $\Lambda$  model. Both semiparametric models estimate a similar value of 0.7 for the precision parameter  $\alpha$ .

Figure 3 and 4 plot the conditional variances and the conditional correlations for the MGARCH-t and the MGARCH-DPM- $\Lambda$  model. These quantities are derived from the posterior mean of the conditional covariances and for the DPM version are computed from (35) for a particular parameter draw. Both models have very similar patterns in their conditional moments through time. Plots for the MGARCH-DPM specification

are very similar.

Finally, the Bayes factor in favor of the MGARCH-t model versus the MGARCH-DPM- $\Lambda$  model were computed for both the equity and foreign exchange data following Section 3.2. For a range of degree of freedom parameters ( $\nu = 6, 8, 12, 30$ ) the Bayes factor is 0. The support for  $u$  is always confined to values below 0.7 as is evident in the posterior density of  $u$  plotted in Figure 5 for the equity data. This indicates the equity data strongly supports a nonparametric specification with only a few (as in ten, more or less), unique covariances,  $B_j$ , as opposed to the MGARCH-t model's 2263 unique covariances.

## 5.2 Out-of-Sample Forecasts

In this section the forecast precision of the models is compared by predictive likelihoods. This measures the overall accuracy of density forecasts from each model and is equivalent to a log-scoring rule (see Winkler (1969)). In each case a model is estimated with data  $Y_{t-1}$  and then the predictive likelihood is estimated for  $y_t$  given  $Y_{t-1}$  following Sections 3.1 and 4.1. After this the data is incremented to  $Y_t$  and the model is re-estimated and the predictive likelihood is computed for  $y_{t+1}$  given  $Y_t$ . This is repeated for all the out-of-sample period  $\tau_1 \leq t \leq \tau_2$  and the log-predictive likelihood  $\sum_{t=\tau_1}^{\tau_2} \log p(y_t|Y_{t-1}, M)$ , where  $M = \text{MGARCH-t, MGARCH-DPM} - \Lambda, \text{MGARCH-DPM}$ , is reported in Table 3 while portfolio results are in Table 4. From this, log-predictive Bayes factors for comparison of two models can be computed from the values in Tables 3 and 4 by subtracting two entries from the same column.

For equity the semiparametric models offer a substantial improvement over the parametric model in prediction. The log-predictive Bayes factor in favour of the MGARCH-DPM- $\Lambda$  over the MGARCH-t is 14.64. For the FX returns there is really no forecast gain compared to the MGARCH-t model as the semiparametric models match the parametric model.

In moving from the MGARCH-DPM- $\Lambda$  model to the more general MGARCH-DPM parametrization there are further gains in prediction accuracy. The log-Bayes factor is 2.63 (equity) in favor of the MGARCH-DPM. Since there is no loss of precision for the FX data and substantial improvement for the equity the MGARCH-DPM is our preferred model. By allowing mixing over the mean vector this model can capture general forms of asymmetry in the return density.

Why do the semiparametric models perform well? Some explanation of this can be

found in Figure 6. This figure reports several features of model comparison using the equity data. The top panel shows the period-by-period difference in the log-predictive likelihoods. A positive (negative) value is in favor of the MGARCH-DPM- $\Lambda$  (MGARCH-t) model. The next panel is the cumulative values from the previous panel. The last panel is the average absolute value of returns for each date of the out-of-sample period. It provides a rough estimate of average volatility.

From Figure 6 we can see three influential observations that are in favor of the MGARCH-DPM- $\Lambda$ . They appear to be high volatility episodes. On the other hand, the second panel shows there to be regular ongoing forecast improvements from the semiparametric model. Even at the end of the sample, which has lower average volatility, the evidence continues to strengthen for the semiparametric model.

In a longer out-of-sample period, Figure 7 displays the sequential estimates of the posterior mean of  $m|Y_t$ , the number of clusters in the DPM mixture, along with average volatility. During the high volatility of the financial crisis the MGARCH-DPM- $\Lambda$  goes from using about 9 components in the mixture to around 10. Towards the end of the sample period the number of clusters begins to decrease. This is a flexible feature of the DPM mixture model. If the future is unlike the past the model can introduce new components with new parameters into the mixture to accommodate new structure in the distribution. This is something that is not possible with the t-distribution or a finite mixture model as the number of components is fixed.

Figure 8 provides a similar comparison between the two semiparametric models. The last panel plots the market return. The MGARCH-DPM model provides regular gains over the whole sample and is not confined to any particular data points such as outliers. Some of the early gains in the model coincide with the negative returns in the market. Recall that the main difference in these models is the ability of the MGARCH-DPM specification to capture asymmetry in the return distribution.

Next we consider results for various portfolios of equity for our best model MGARCH-DPM along with the Student-t version. Table 4 displays the cumulative log-predictive likelihoods for the full support of the distribution along with the conditional (tail) log-predictive likelihood for portfolio returns less than  $-1.0$ . The final column of this table lists the portfolio weights. The MGARCH-DPM specification gives robust gains for different portfolios for the whole distribution. The log-Bayes factor is on the order of 4 for each portfolio. On the other hand, the tail predictive likelihoods are almost identical between the models with a slight edge for the DPM version. In contrast to the previous results, in which the same tail behaviour of the Student-t distribution is imposed in

3 dimensions and worked poorly, the MGARCH-t only has to match the tail in one dimension for the univariate portfolio. As such it does much better in this situation.

As an example of the breakdown of these results, Figure 9 plots the portfolio return versus the difference in the log-predictive likelihoods for the MGARCH-DPM and the MGARCH-t, for the third portfolio in Table 4. A positive (negative) value on the y-axis is in favor of the MGARCH-DPM (MGARCH-t). Overall the models are very similar. The DPM version is better in the range  $[-5;-1]$  and  $[1;5]$  while the Student-t model is better around 0 and in the extreme tails.

In summary, the semiparametric models are at least as good as the parametric model with t-innovations and often significantly better in terms of density forecasts. This has been demonstrated for three assets and portfolios. We now turn to a more demanding application of the model.

### 5.3 Ten Asset Example

In this section we show that our approach can be used for a larger portfolio of assets. We consider the previous models and propose some simplifications, that in some cases, result in better forecast performance. Using the same sample period 2001/01/02 – 2009/12/31 (2263 observations) returns from CRSP with company ticker symbols: HWP, AAPL, IBM, XOM, JPM, MET, GE, FDX, DLM, and WHR are used.

In this case, with  $k = 10$  assets the MGARCH model of (8) now has 75 ( $k(k + 1)/2 + 2k$ ) parameters as well as the parameters associated with the DPM component of the model. Clearly, computational considerations begin to play an important role. In order to compare the most flexible specification in Section 4 with some restricted versions we employ OpenMP<sup>4</sup> for parallel programming on an Intel quadcore processor. The main bottleneck is associated with computing the loglikelihood function in (32). GARCH models have a path dependence in that the covariance matrix at time  $t$  affects all future covariance matrices. This structure does not readily admit standard parallel programming methods. To do so, we first run through the MGARCH recursion to obtain  $\{H_t\}_{t=1}^T$  and then compute the conditional covariances (35). After this the Cholesky decomposition of each of the covariance terms, needed for density evaluation, can be calculated independently among the 4 processors. Although additional gains are possible, this is sufficient to make the computation time practical. For instance, the full model in Section 4, given an estimate of  $V$ , the covariance of the proposal density (4)

---

<sup>4</sup><http://openmp.org/wp/>

used to sample the MGARCH parameters, it only takes a few minutes to obtain 8000 draws from the full posterior distribution of which the last 5000 are used for inference.

We consider several restricted models and some require modification to the MCMC sampler. Each semiparametric model in this section uses the slice sampler for estimation. First we restrict all mixture mean vectors in the DPM to zero, in other words,  $\mu_t \equiv 0$ , and secondly, the mixture precision matrices,  $B_j$ , are diagonal. This forces all nondiagonal elements to be zero but allows separate scaling for each element of  $x_t$ . The Wishart prior for  $B_j$  is reduced to the Gamma prior

$$B_{j,ii} \sim \text{Gamma} \left( \frac{\nu + k - 1}{2}, \frac{P_{ii}^{-1}}{2} \right), \quad i = 1, \dots, k, \quad (67)$$

which implies  $E[B_{j,ii}] = 1$  if  $P_{ii} = 1/(\nu + k - 1)$ . For each active cluster,  $j = 1, \dots, m$ , the Gibbs update step becomes

$$B_{j,ii} | \{x_t : s_t = j\} \sim \text{Gamma} \left( \frac{n_j + \nu + k - 1}{2}, \frac{\sum_{t:s_t=j} x_{ti}^2 + P_{ii}^{-1}}{2} \right), \quad (68)$$

$i = 1, \dots, k$ , while  $B_{j,il} = 0, i \neq l$ .

A further simplification is to restrict  $B_j = b_j I_k$  where  $b_j$  is a scalar. This scales the whole matrix  $H_t$  up or down by a fixed amount in each cluster of the DPM. In this case with prior

$$b_j \sim \text{Gamma} \left( \frac{\nu + k - 1}{2}, \frac{p^{-1}}{2} \right), \quad (69)$$

and  $p = 1/(\nu + k - 1)$  the Gibbs step in (68) becomes one draw from

$$b_j | \{x_t : s_t = j\} \sim \text{Gamma} \left( \frac{n_j k + \nu + k - 1}{2}, \frac{\sum_{t:s_t=j} x'_t x_t + p^{-1}}{2} \right)$$

for  $j = 1, \dots, m$ .

Table 5 reports the predictive likelihoods for the same sample period as before but for the 10 assets. The same priors are used as in the previous examples and any changes are listed in the center column of Table 5. The first model is the parametric MGARCH-t model with a cumulative log-predictive likelihood of  $-5549.69$ . The next specification is our most flexible semiparametric model of Section 4, the MGARCH-DPM model. The average number of clusters  $m$  with this model is around 25 but it often has MCMC draws well in excess of 50. With a cumulative log-predictive likelihood of  $-5551.94$ , this model is slightly worse than the MGARCH-t model.

The third through fifth row of Table 5 impose restrictions on the MGARCH-DPM model in one form or another. In general, the tighter prior  $\text{Gamma}(0.5, 12)$  for the precision parameter  $\alpha$  results in the average number of active clusters being around 20. It also delivers better performance in the form of the cumulative log-predictive likelihood being approximately  $-5545$ .

Imposing a zero mean vector  $\xi_t = 0$  further improves the density forecasts of returns. The best specification features this restriction as well as a diagonal  $B_j$  matrix. This restricted semiparametric MGARCH model has a log-predictive Bayes factor of 28.1 against the MGARCH-t – strong evidence in favor of the semiparametric model. The final model restricts the cluster covariance to a scaled identity matrix and performs poorly with the smallest cumulative log-predictive likelihood of  $-5556.25$ . Except for this last specification all of the semiparametric models are competitive with the parametric model and often give significantly more accurate forecasts.

## 6 Conclusion

This paper proposes a Bayesian nonparametric modeling approach for the return innovations in multivariate GARCH models. The approach of this paper opens the door to semiparametric modeling in a wide variety of different multivariate GARCH models. An infinite mixture of multivariate normals is given a flexible Dirichlet process prior. We discuss conjugate methods that allow for scale mixtures and nonconjugate methods which provide mixing over both the location and scale matrix of the normal components. MCMC methods are introduced for posterior simulation and computation of the predictive density. Bayes factors and density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach. The paper concludes with an application to 10 assets. We introduce several restricted versions of the model that reduce computational costs and lead to better out-of-sample forecast performance. The best semiparametric model continues to have significant improvements in density forecasts of returns beyond the MGARCH model with Student-t innovations.

## References

- Ausín, C., Galeano, P. & Ghosh, P. (2010), A semiparametric bayesian approach to the analysis of financial time series with applications to value at risk estimation, Statistics and econometrics working papers, Universidad Carlos III, Departamento de Estadística y Econometría.
- Bali, T. G. & Engle, R. F. (2010), ‘The intertemporal capital asset pricing model with dynamic conditional correlation’, *Journal of Monetary Economics* **57**, 337–390.
- Bauwens, L., Hafner, C. & Rombouts, J. (2007), ‘Multivariate mixed normal conditional heteroskedasticity’, *Computational Statistics & Data Analysis* **51**(7), 3551 – 3566.
- 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**(3), 346–354.
- Bollerslev, T., Engle, R. & Wooldridge, J. (1988), ‘A capital asset pricing model with time varying covariances’, *Journal of Political Economy* **96**, 116–131.
- Buraschi, A., Porchia, P. & Trojani, F. (2010), ‘Correlation risk and optimal portfolio choice’, *Journal of Finance* **65**, 293–420.
- Dellaportas, P. & Vrontos, I. D. (2007), ‘Modelling volatility asymmetries: a bayesian analysis of a class of tree structured multivariate garch models’, *Econometrics Journal* **10**(3), 503–520.
- Diamantopoulos, K. & Vrontos, I. D. (2010), ‘A student-t full factor multivariate garch model’, *Computational Economics* **35**(1), 63–83.
- Dickey, J. J. (1971), ‘The weighted likelihood ratio, lineary hypotheses on normal location parameters’, *The Annals of Mathematical Statistics* .
- Diks, C., Panchenko, V. & van Dijk, D. (2011), ‘Likelihood-based scoring rules for comparing density forecasts in tails’, *Journal of Econometrics* **163**(2), 215 – 230.
- Ding, Z. & Engle, R. (2001), ‘Large scale conditional covariance matrix modeling estimation and testing’, *Academia Economic Papers* **29**(2), 157–184.
- Escobar, M. D. & West, M. (1995), ‘Bayesian density estimation and inference using mixtures’, *Journal of the American Statistical Association* **90**(430), 577–588.



- Ferguson, T. (1973), ‘A Bayesian analysis of some nonparametric problems’, *The Annals of Statistics* **1**(2), 209–230.
- 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 & Economic Statistics* **21**(4), 532–546.
- 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 & Economic Statistics* **28**(4), 559–571.
- Hafner, C. M. & Rombouts, J. (2007), ‘Semiparametric multivariate volatility models’, *Econometric Theory* **23**, 251–280.
- Harvey, A., Ruiz, E. & Sentana, E. (1992), ‘Unobserved component time series models with arch disturbances’, *Journal of Econometrics* **52**(1-2), 129–157.
- Hudson, B. & Gerlach, R. (2008), ‘A bayesian approach to relaxing parameter restrictions in multivariate garch models’, *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research* **17**(3), 606–627.
- Ishwaran, H. & Zarepour, M. (2000), ‘Markov chain monte carlo in approximate dirichlet and beta two-parameter process hierarchical models’, *Biometrika* **87**(2), pp. 371–390.
- Jensen, M. J. & Maheu, J. M. (2010), ‘Bayesian semiparametric stochastic volatility modeling’, *Journal of Econometrics* **157**(2), 306 – 316.
- Jensen, M. J. & Maheu, J. M. (2012), Estimating a semiparametric asymmetric stochastic volatility model with a dirichlet process mixture. Department of Economics, Working Paper 453, University of Toronto.
- Kalli, M., Griffin, J. & Walker, S. (2011), ‘Slice sampling mixture models’, *Statistics and Computing* **21**, 93–105.
- Laurent, S., Bauwens, L. & Rombouts, J. V. K. (2006), ‘Multivariate garch models: a survey’, *Journal of Applied Econometrics* **21**(1), 79–109.

- Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), ‘Flexible multivariate garch modeling with an application to international stock markets’, *The Review of Economics and Statistics* **85**(3), 735–747.
- Long, X., Su, L. & Ullah, A. (2011), ‘Estimation and forecasting of dynamic conditional covariance: A semiparametric multivariate model’, *Journal of Business and Economic Statistics* **29**(1), 109–125.
- Osiewalski, J. & Pipien, M. (2004), ‘Bayesian comparison of bivariate arch-type models for the main exchange rates in poland’, *Journal of Econometrics* **123**(2), 371 – 391.
- Papaspiliopoulos, O. (2008), A note on posterior sampling from dirichlet mixture models. manuscript, Department of Economics, Universitat Pompeu Fabra.
- Richardson, M. & Smith, T. (1993), ‘A test of multivariate normality of stock returns’, *Journal of Business* **66**, 295–321.
- 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’, Springer-Verlag, pp. 201–226.
- Walker, S. G. (2007), ‘Sampling the dirichlet mixture model with slices’, *Communications in Statistics – Simulation and Computation* **36**, 45–54.
- West, M., Muller, P. & Escobar, M. (1994), Hierarchical priors and mixture models with applications in regression and density estimation, in P. R. Freeman & A. F. Smith, eds, ‘Aspects of Uncertainty’, John Wiley.
- Winkler, R. L. (1969), ‘Scoring rules and the evaluation of probability assessors’, *Journal of the American Statistical Association* **64**, 1073–1078.

Table 1: Summary Statistics

	Mean	Variance	Skewness	Kurtosis	Sample Correlation		
IBM	0.0389	3.0940	0.5253	9.4160	1		
VW	0.0159	1.9276	-0.0010	11.0670	0.6889	1	
HP	0.0850	8.4305	-0.1016	9.9779	0.3271	0.5758	1
<hr/>							
Euro-USD	-0.0044	0.4249	-0.2092	5.4101	1		
UK-USD	0.0030	0.3768	0.1609	7.8475	0.6802	1	
JPY-USD	-0.0069	0.4697	-0.4093	5.7904	0.2712	0.1412	1

---

Equity is the daily return on IBM, value-weighted CRSP portfolio (VW), HP obtained from CRSP, 2001/01/02 – 2009/12/31 (2263 observations). FX is log-returns on Euro-USD, UK-USD, JPY-USD, 1999/01/05 – 2010/04/19 (2834 observations).

Table 2: Full Sample Estimates, CRSP Data

	MGARCH-t		MGARCH-DPM- $\Lambda$		MGARCH-DPM	
	Mean	0.95 DI	Mean	0.95 DI	Mean	0.95 DI
$g_{11}$	0.1017	( 0.0748, 0.1288)	0.1200	(0.0849, 0.1545)	0.1255	( 0.0949, 0.1610)
$g_{21}$	0.0581	( 0.0362, 0.0856)	0.0317	(-0.0152, 0.0727)	0.0268	(-0.0068, 0.0575)
$g_{31}$	0.0418	( 0.0025, 0.0816)	0.0084	(-0.0741, 0.0988)	0.0207	(-0.0460, 0.0890)
$g_{22}$	0.0623	( 0.0453, 0.0795)	0.0732	( 0.0533, 0.0933)	0.0753	( 0.0511, 0.0980)
$g_{32}$	0.0926	( 0.0518, 0.1357)	0.0996	( 0.0377, 0.1723)	0.0088	( 0.0093, 0.1758)
$g_{33}$	0.1633	( 0.1075, 0.2147)	0.1214	( 0.0428, 0.1790)	0.1281	( 0.0068, 0.1877)
$\gamma_{11}$	0.1534	( 0.1341, 0.1735)	0.1914	( 0.1626, 0.2198)	0.1951	( 0.1625, 0.2366)
$\gamma_{12}$	0.1883	( 0.1619, 0.2165)	0.1936	( 0.1615, 0.2228)	0.1930	( 0.1602, 0.2231)
$\gamma_{13}$	0.1636	( 0.1428, 0.1865)	0.1570	( 0.1318, 0.1839)	0.1563	( 0.1307, 0.1828)
$\gamma_{21}$	0.9793	( 0.9737, 0.9838)	0.9777	( 0.9721, 0.9823)	0.9769	( 0.9698, 0.9818)
$\gamma_{22}$	0.9721	( 0.9620, 0.9796)	0.9649	( 0.9577, 0.9719)	0.9663	( 0.9574, 0.9745)
$\gamma_{23}$	0.9787	( 0.9731, 0.9836)	0.9746	( 0.9687, 0.9796)	0.9754	( 0.9691, 0.9805)
$\psi$	7.7147	( 6.5995, 8.9348)				
$\alpha$			0.6784	( 0.2391, 1.3544)	0.7159	(0.2351, 1.4586)
$m$			9.4901	( 5.0000, 16.000)	10.0994	( 5.0000, 17.0000)

This table displays the posterior mean and the 0.95 density intervals (DI) of model parameters. Data is daily return on IBM, vwretd, HP obtained from CRSP, 2001/01/02 – 2009/12/31 (2263 observations).  $(g_{11}, \dots, g_{33})' = \text{vech}(\Gamma_0^{1/2})$ .

Table 3: Cumulative Log-Predictive Likelihoods

	Equity	FX
MGARCH-t	-1468.26	-1096.47
MGARCH-DPM- $\Lambda$	-1453.62	-1096.63
MGARCH-DPM	-1450.97	-1096.88
Observations	262	333

FX is log-returns on Euro-USD, UK-USD, JPY-USD, 1999/01/05 – 2010/04/19 (2834 observations). Predictive likelihoods are computed for the final 333 observations at the end of the sample. Equity is the daily return on IBM, vwretd, HP obtained from CRSP, 2001/01/02 – 2009/12/31 (2263 observations). Predictive likelihoods are computed for the final 262 observations at the end of the sample.

Table 4: Cumulative Log-Predictive Likelihoods for Equity Portfolios

	Log-Predictive Likelihoods	Tail Log-Predictive Likelihoods	Portfolio Weight
MGARCH-t	-548.29	-89.00	1/3,1/3,1/3
MGARCH-DPM	-544.77	-88.57	1/3,1/3,1/3
MGARCH-t	-506.76	-66.01	3/5,1/5,1/5
MGARCH-DPM	-504.87	-65.90	3/5,1/5,1/5
MGARCH-t	-522.43	-75.76	1/5,3/5,1/5
MGARCH-DPM	-518.84	-75.41	1/5,3/5,1/5
MGARCH-t	-616.13	-119.80	1/5,1/5,3/5
MGARCH-DPM	-612.33	-119.04	1/5,1/5,3/5
Observations	262	62	

The portfolio comprises IBM, vwret, HP obtained from CRSP, 2001/01/02 – 2009/12/31 (2263 observations) and using portfolio weights in the last column. Predictive likelihoods are computed for the final 262 observations at the end of the sample while the tail log-predictive likelihood is  $p(y_{t+1}^p | y_{t+1}^p < \eta, Y_t)$  for  $\eta = -1.0$ .

Table 5: Cumulative Log-Predictive Likelihoods 10 Assets

Model	Prior/Restriction	
MGARCH-t		-5549.69
MGARCH-DPM		-5551.94
MGARCH-DPM	$\alpha \sim \text{Gamma}(0.5, 12), \xi_t \sim N(0, 0.1)$	-5545.58
MGARCH-DPM	$\alpha \sim \text{Gamma}(0.5, 12), \xi_t \sim N(0, 0.1), \nu = 15$	-5544.88
MGARCH-DPM	$\xi_t = 0$	-5538.55
MGARCH-DPM	$\xi_t = 0, B_t = \text{diag}(b_1, \dots, b_k)$	-5521.59
MGARCH-DPM	$\xi_t = 0, B_t = bI_k$	-5556.29
Observations		262

Equity is the daily return on the following stocks with symbols: HWP, AAPL, IBM, XOM, JPM, MET, GE, FDX, DLM, and WHR, 2001/01/02 – 2009/12/31 (2263 observations). Predictive likelihoods are computed for the final 262 observations at the end of the sample. The benchmark prior is listed in Section 5.1 and any deviations from this are noted in the center column of the table.

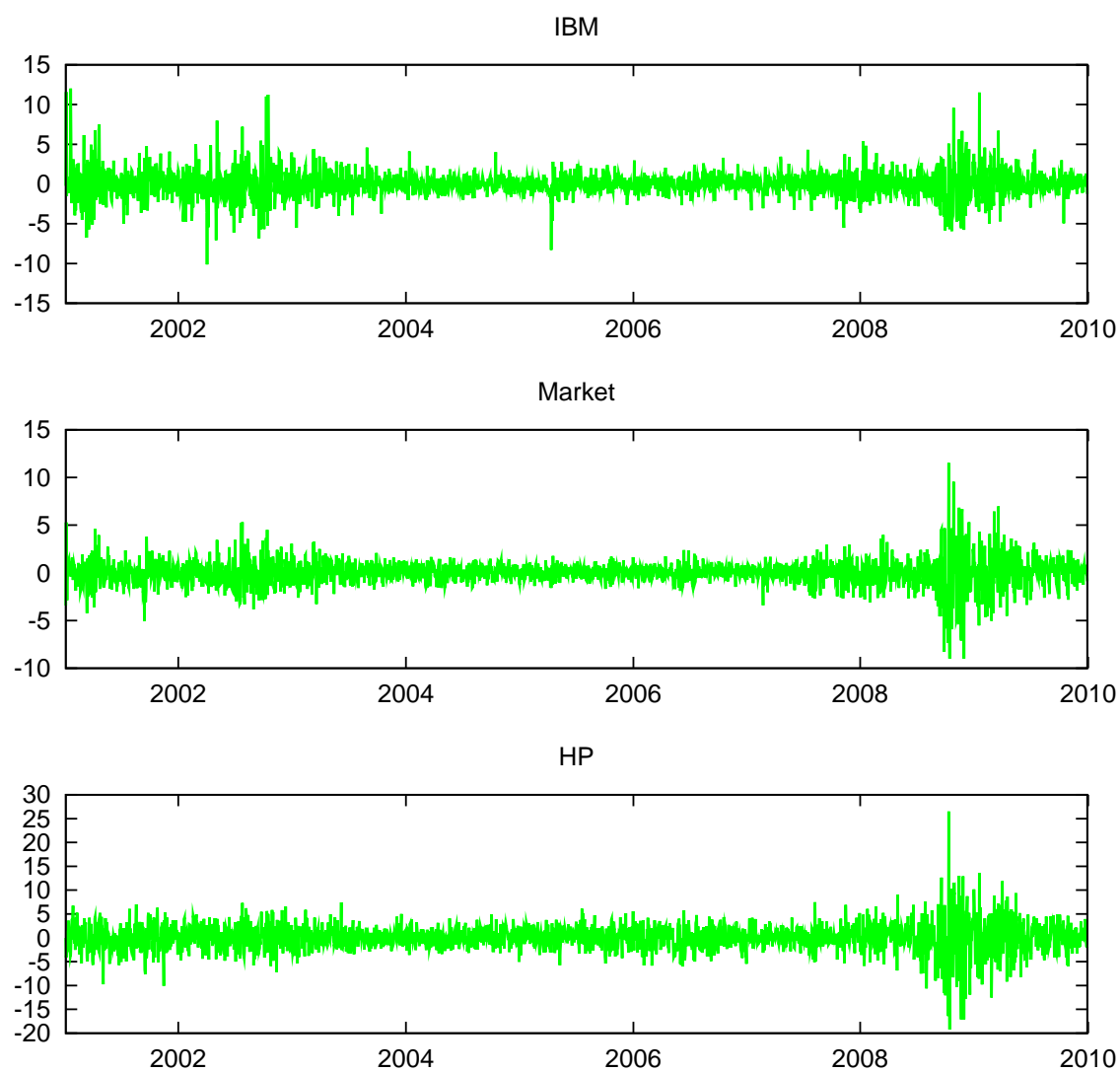


Figure 1: Daily equity returns for IBM, the market, value-weighted CRSP portfolio, and Hewlett and Packard.

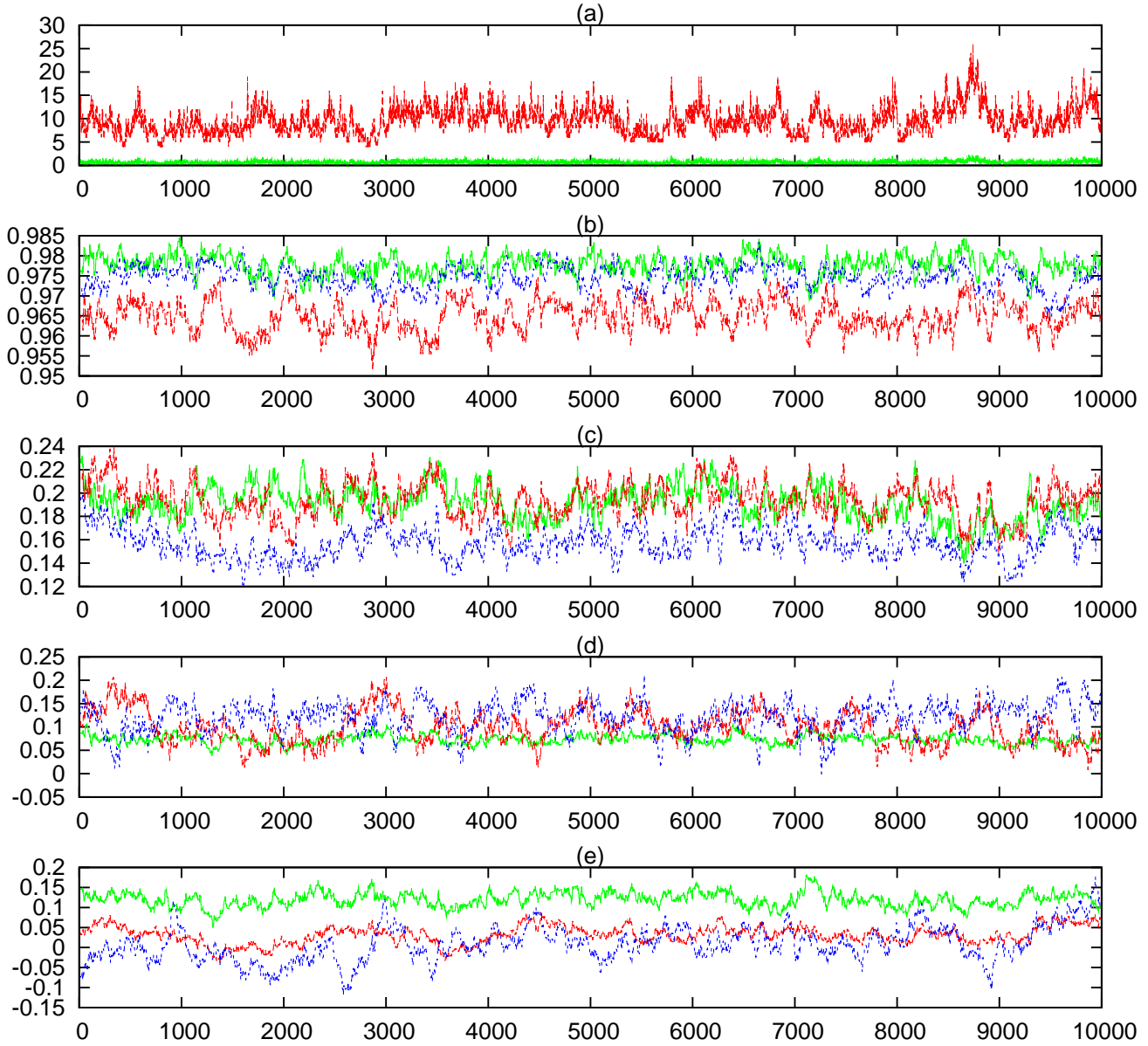


Figure 2: Posterior draws of the MGARCH-DPM- $\Lambda$  parameters  $\alpha$  and  $m$  in panel (a),  $\gamma_2$  in panel (b),  $\gamma_1$  in panel (c), the first three elements of  $vech(\Gamma_0^{1/2})$  in panel (d), and the last three elements of  $vech(\Gamma_0^{1/2})$  in panel (e).

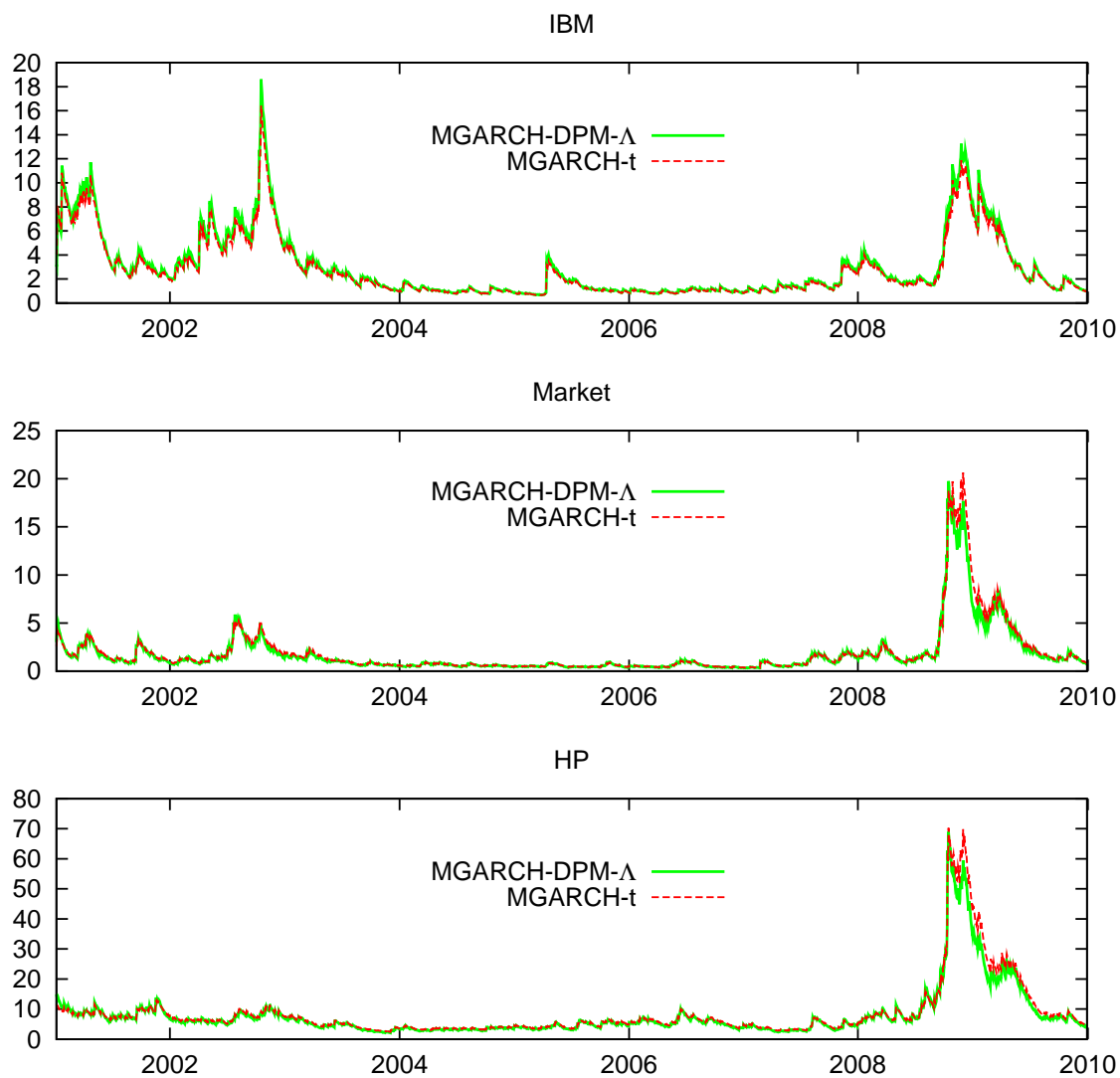


Figure 3: Posterior mean of the conditional variances,  $\text{Var}(y_t|Y_T)$ .



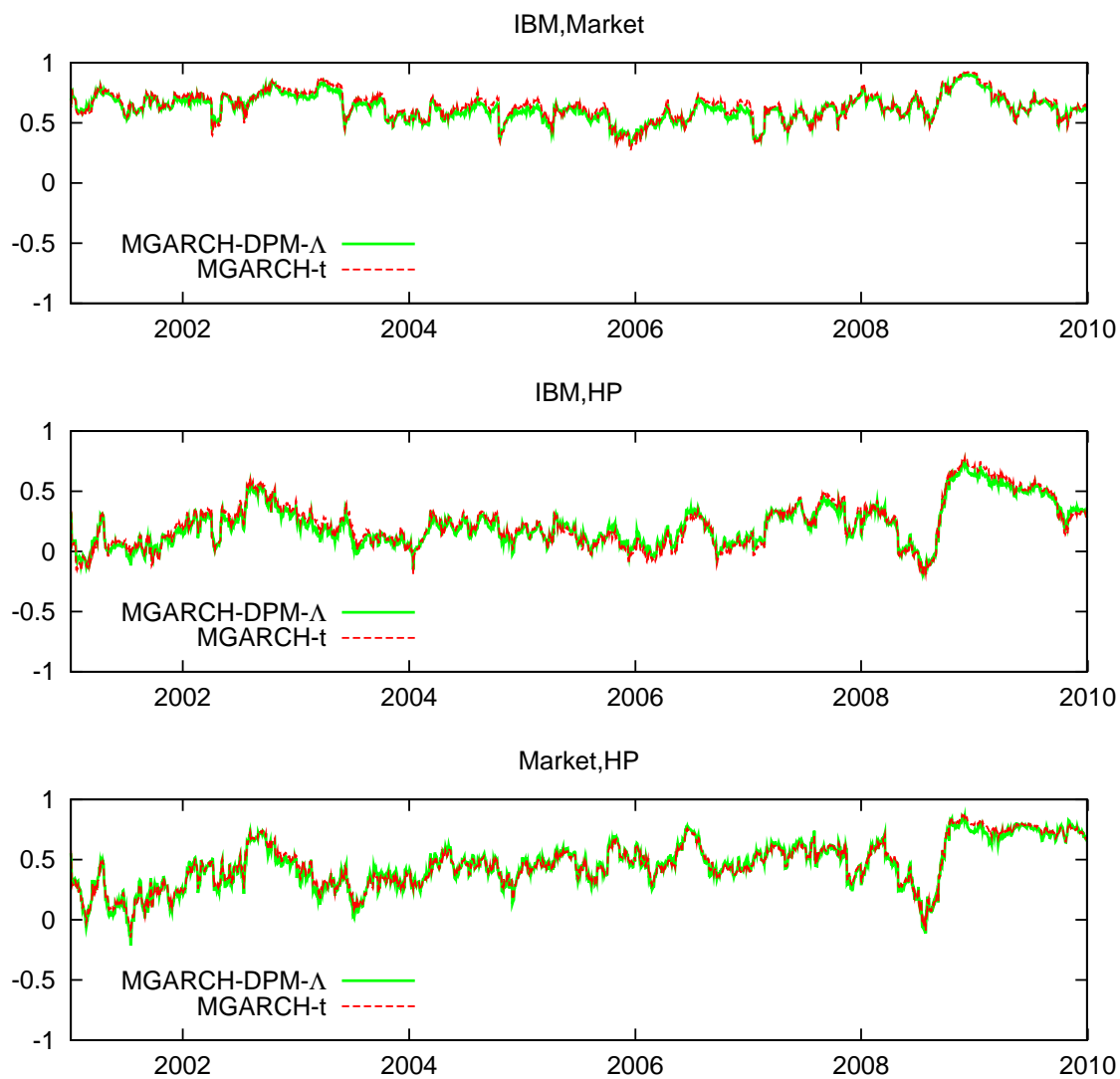


Figure 4: Posterior mean of the conditional correlations,  $\text{Corr}(y_t|Y_T)$ .

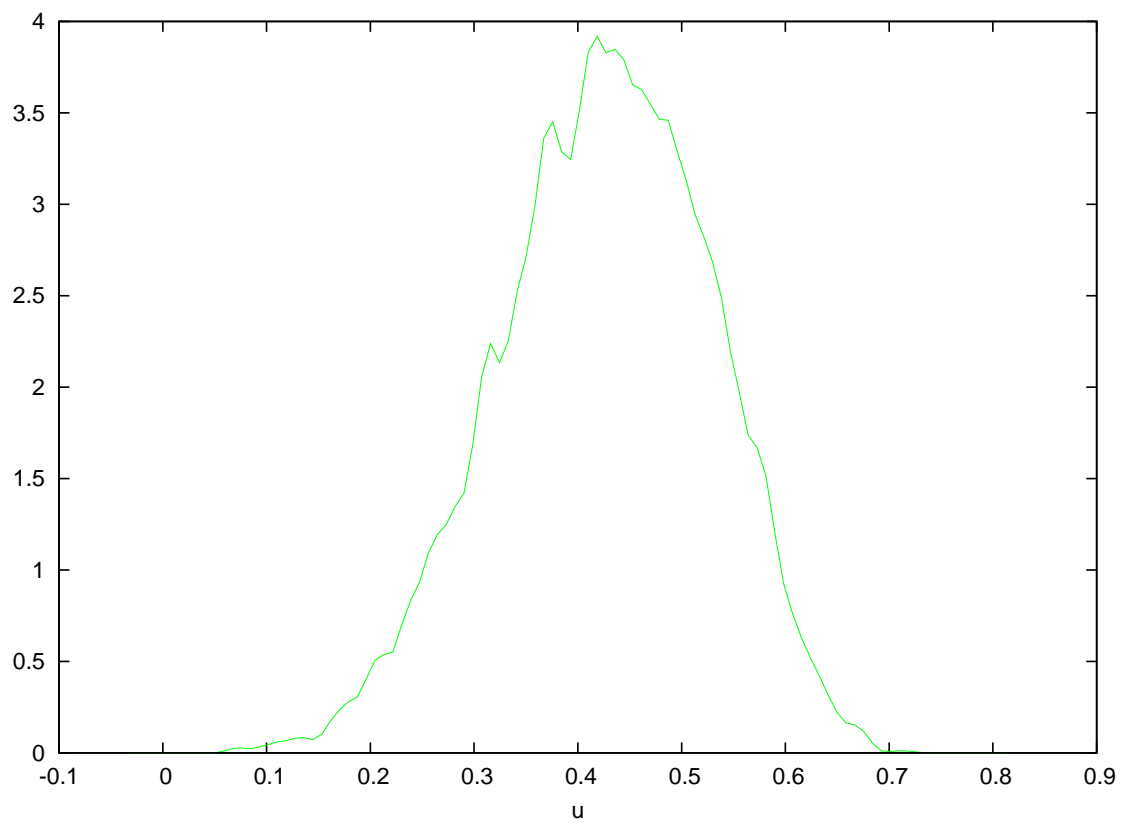


Figure 5: Posterior density of  $u \equiv \alpha/(1 + \alpha)$ . The MGARCH-t is nested in the MGARCH-DPM- $\Lambda$  model when  $u = 1$ .

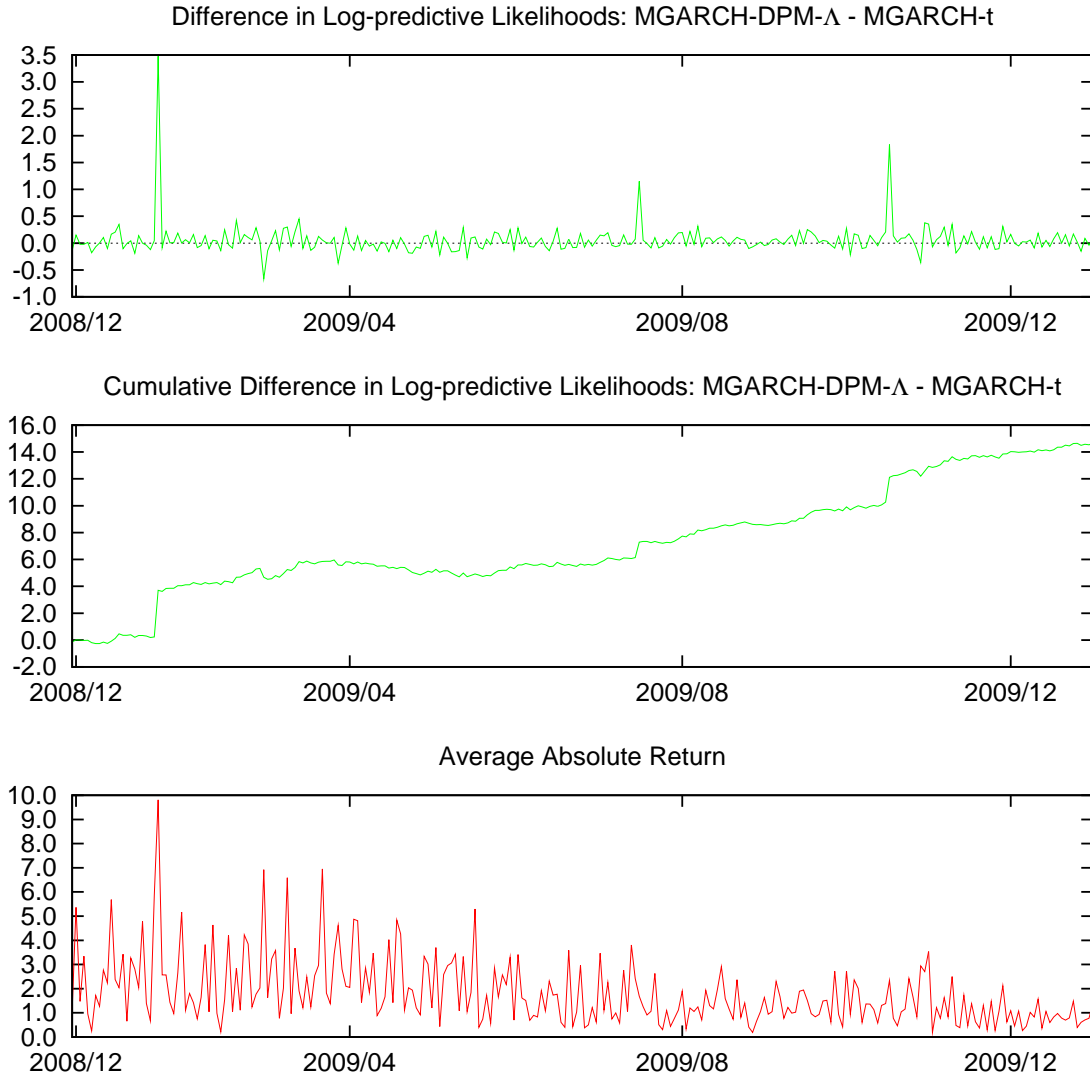


Figure 6: Model comparison between the MGARCH-DPM- $\Lambda$  and MGARCH-t models for equities using  $\sum_{t=2002}^{2263} [\log p(y_t|Y_{t-1}, \text{MGARCH-DPM} - \Lambda) - \log p(y_t|Y_{t-1}, \text{MGARCH-t})]$ .

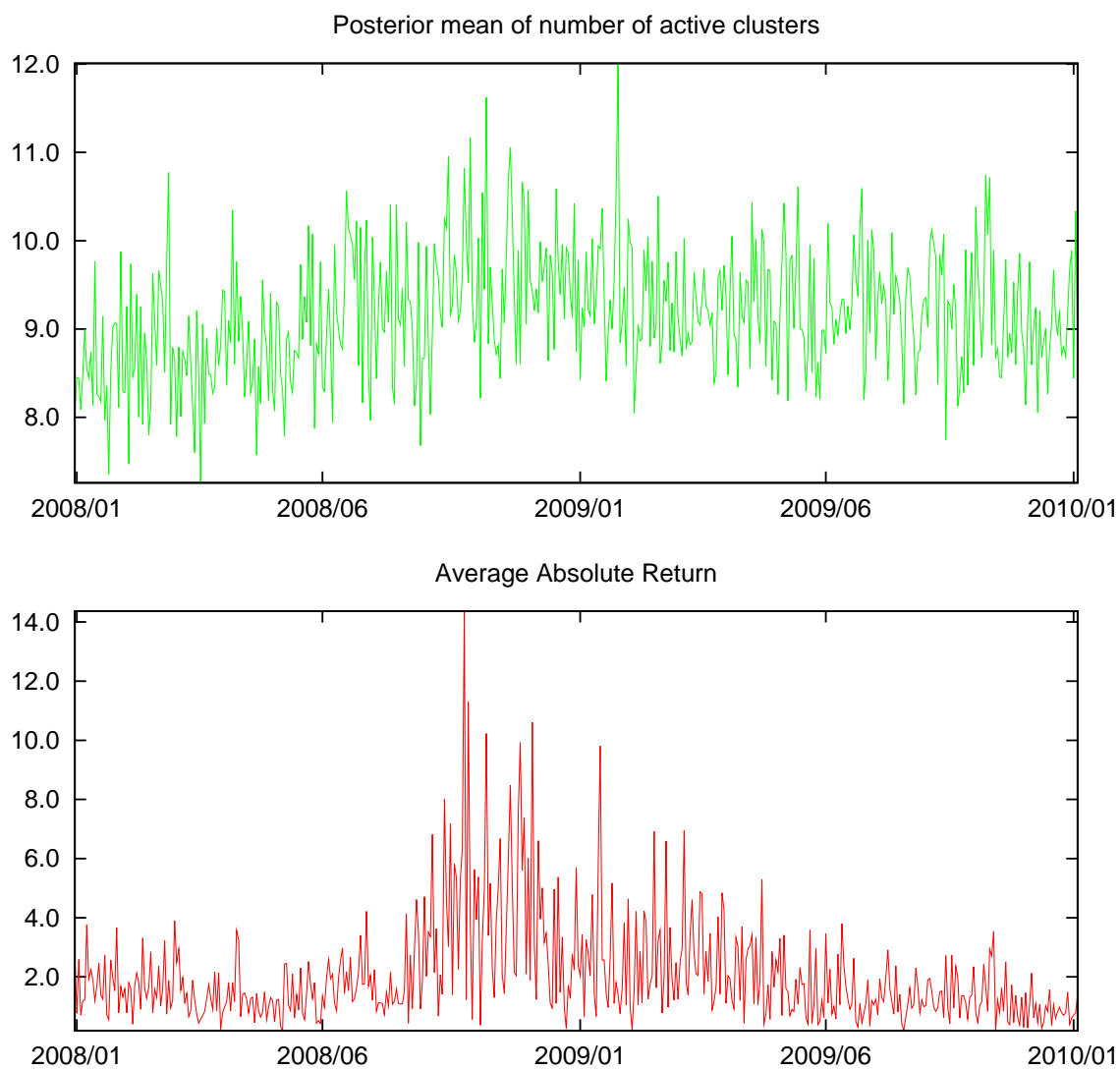


Figure 7: Sequential estimates of the mean of  $p(m|Y_t)$  from January 2, 2008 to December 31, 2009 with the MGARCH-DPM- $\Lambda$  model.

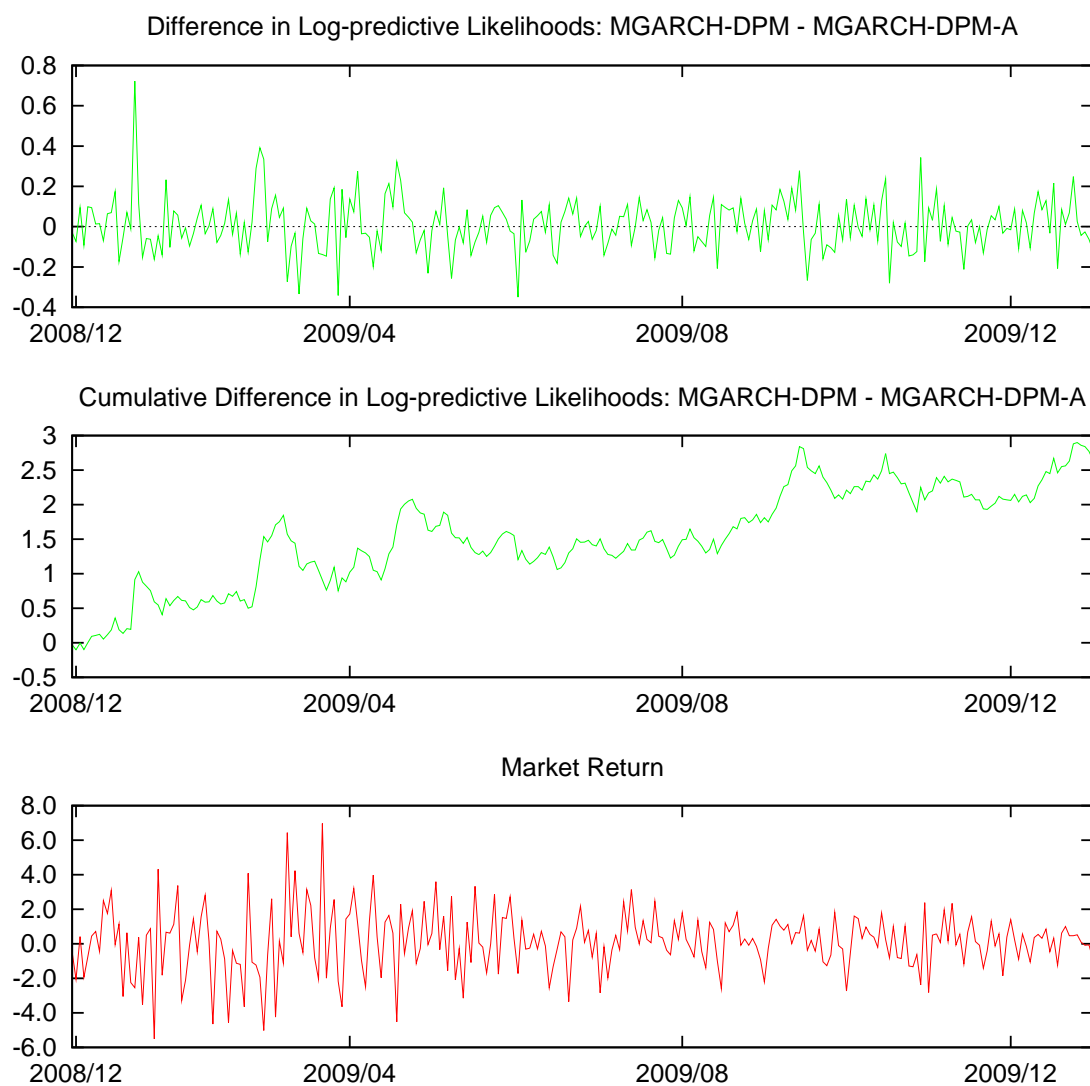


Figure 8: Model comparison between the semiparametric MGARCH-DPM and MGARCH-DPM- $\Lambda$  models for equities using  $\sum_{t=2002}^{2263} [\log p(y_t|Y_{t-1}, \text{MGARCH-DPM}) - \log p(y_t|Y_{t-1}, \text{MGARCH-DPM} - \Lambda)]$ .

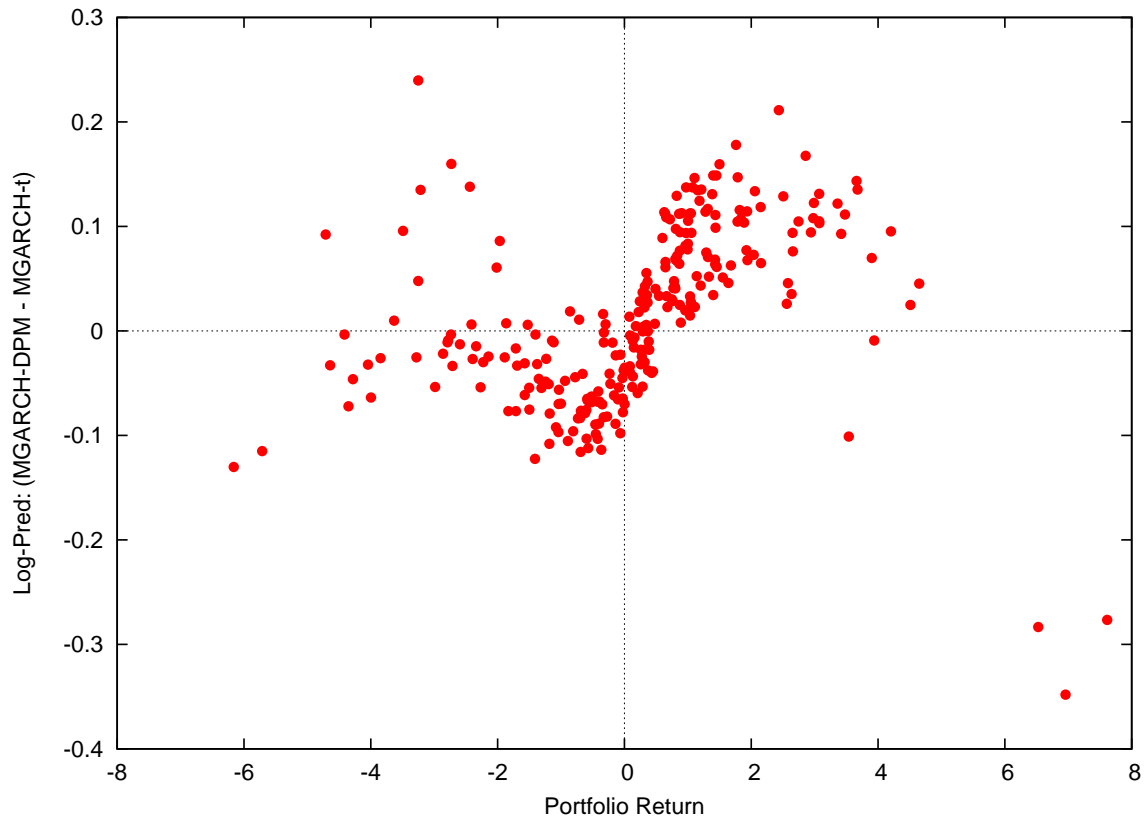


Figure 9: Portfolio return  $y_t^p$  versus  $[\log p(y_t^p|Y_{t-1}, \text{MGARCH-DPM}) - \log p(y_t^p|Y_{t-1}, \text{MGARCH-t})]$ .