Bayesian semiparametric multivariate GARCH modeling*

Mark J. Jensen† and John M. Maheu‡

preliminary
October 2010

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. Density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach.

*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.

†Federal Reserve Bank of Atlanta, Mark.Jensen@atl.frb.org

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

1 Introduction

This paper proposes a Bayesian nonparametric modeling approach for the return innovations 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. 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. Several MCMC methods are introduced for posterior simulation and computation of the predictive density. Density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach.

Modeling comovements in asset returns has been a central topic in empirical finance for several decades. The multivariate GARCH (MGARCH) model continues to be the main class of models used in empirical application. Recent surveys of this literature can be found in Laurent et al. (2006) and Silvennoinen & Tersvirta (2007). The importance of considering other innovation distributions beyond a normal has been recognized by Diamantopoulos & Vrontos (2010), Harvey et al. (1992) and Fiorentini et al. (2003) which adopt a Student-t density. Bauwens & Laurent (2005) extend this to a multivariate skew-Student distribution. Finite mixtures are treated in Bauwens et al. (2007). Each of these extensions improves the model fit but all remain essentially parametric approaches. Classical semiparametric approaches to MGARCH estimation are considered by Long & Ullah (2005) and Hafner & Rombouts (2007). They employ a two-step estimation strategy.

It is unclear what the finite sample performance of MGARCH model estimators are and the effect of two-step estimation. On the other hand, an open issue is the asymptotic normality of maximum likelihood estimators (point 9 of Bauwens et al. (2007)). Bayesian methods are attractive in that estimation is conceptually straightforward and avoids these problems. However, existing Bayesian estimation by Dellaportas & Vrontos (2007), Hudson & Gerlach (2008) and Osiewalski & Pipien (2004) are all based on parametric assumptions. The purpose of this paper is to extend this work to a 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 innovation distribution. The predictive density integrates out uncertainty from the conditional covariance as well as uncertainty about the return distribution.

Our work is closely related to the semiparametric univariate volatility models of Jensen & Maheu (2010) and Ausín et al. (2010) which are based on the Dirichlet process prior for the conditional distribution of returns. This is a flexible prior of a countably infinite mixture of distributions. Our first approach makes use of the Polya urn sampling methods from Escobar & West (1995) and West et al. (1994) for the Dirichlet process mixture (DPM) model to nonparametrically estimate the MGARCH model's return distribution. Each of the components of the mixture have a different covariance matrix and permits conjugate posterior sampling for the DPM portion of the model.

In order to allow different location vectors in the mixture components, which will be important to capture potential asymmetry in the return distribution, an approximation of the infinite ordered DPM prior's stick breaking representation is used. Posterior sampling methods are straightforward (Ishwaran & Zarepour (2000, 2002)) and allow inference on the return distribution itself which is not possible with the Gibbs based Polya urn sampling approach since the latter marginalizes over the 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 conditional correlations are likewise similar. In predictive likelihood comparisons the semi-parametric model is at least as good as the parametric alternative and often significantly better. For example, in the equity application the predictive log-Bayes factor in favor of the new model is 14. On average 10 components are used in the mixture to approximate the distribution of returns.

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 more general DPM model that allows the mean of the components in the infinite mixture to be non-zero. The data is discussed in Section 5. Estimation results and model comparison are found in Section 6 while the paper concludes in Section 7.

2 Benchmark Model

In the following let $y_t = (y_{1t}, \ldots, y_{kt})'$ 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

$$
\begin{aligned}
y_t|H_t &\sim St(0, H_t, \psi) && (1)\\
H_t &= \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}' + \Gamma_2 \odot H_{t-1}, && (2)
\end{aligned}
$$

where $St(0, H_t, \psi)$ is a k variate multivariate Student-t distribution with mean vector 0, $k \times k$ scale matrix H_t and $\psi > 2$ 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 Γ_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 γ_1 and γ_2 are k vectors. The symbol $\odot$ denotes the Hadamard product. This model assumes that each conditional second moment H_t is only related to a lag of itself plus an innovation shock $y_{t-1}y_{t-1}'$. The rank one restrictions on Γ_1 and Γ_1 serve to keep the model parsimonious. Several features of this model are discussed in detail in Ledoit et al. (2003).

3 Multivariate GARCH-DPM model

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

$$
\begin{aligned}
y_t|\Lambda_t, H_t &\sim N(0, H_t^{1/2}\Lambda_t^{-1}(H_t^{1/2})'), && (3)\\
H_t &= \Gamma_0 + \Gamma_1 \odot y_{t-1}y_{t-1}' + \Gamma_2 \odot H_{t-1} && (4)\\
\Lambda_t|G &\overset{iid}{\sim} G, && (5)\\
G|G_0, \alpha &\sim \text{DP}(G_0, \alpha), && (6)\\
G_0(\Lambda_t) &\equiv \text{Wishart}_k(P, v + k - 1),\ v \geq 1. && (7)
\end{aligned}
$$

The MGARCH parameters Γ_0, Γ_1 and Γ_2 follow the same structure as in the previous section. (3),(5)-(7) places a nonparametric prior on the random unconditional 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 parameter Λ_t is assumed to be distributed according to an unknown distribution G. G is modeled nonparametrically as a Dirichlet process (DP) prior with base measure $G_0(\Lambda_t)$ and precision parameter α. The base measure is a Wishart density of dimension k with scale matrix P and degrees of freedom $v + k - 1$ 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

$$y_t|H_t \sim \sum_{j=1}^{\infty} V_j f_k\left(y_t|0, H_t^{1/2} D_j^{-1} (H_t^{1/2})'\right) \tag{8}$$

where $f_k\left(y_t|0, H_t^{1/2} D_j^{-1} (H_t^{1/2})'\right)$ 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 $W_j \sim Beta(1, \alpha)$. The mixture parameters D_j^{-1} prior is found in (7).

In this formulation of the DPM prior only the precision matrix Λ_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 $(\Lambda_1, \dots, \Lambda_T)$. Due to the nature of the DPM model some of the Λ_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 Λ_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 $S^{(t)}$ 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 Λ_{-t}. To describe the sampler for $B, S|Y_T, H_1, \dots, H_T$ we first rewrite Equation (3) so that:

$$w_t|\Lambda_t \sim N(0, \Lambda_t^{-1}), \tag{9}$$

where $w_t \equiv H^{-1/2} y_t$.

Draws are now made from $B, S|w$ with the following two step procedure:

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

$$\begin{aligned} \Lambda_t|w_t, \Lambda_{-t}, s^{(t)} \quad &\sim \quad c\frac{\alpha}{\alpha + T - 1} g(w_t)\, G(d\Lambda_t|w_t) \\ &+ \frac{c}{\alpha + T - 1} \sum_{j=1}^{m^{(t)}} n_j^{(t)} f_k(w_t|0, B_j^{-1})\, \delta_{B_j}(\Lambda_t). \end{aligned} \tag{10}$$

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

$$B_j|\{w_t : s_t = j\} \propto \prod_{t:s_t=j} f_k\left(w_t|0, \Lambda_t^{-1}\right) G_0(d\Lambda_t). \tag{11}$$

We require the following derivations.[1]

$$
\begin{aligned}
G(d\Lambda_t|w_t) &\equiv \frac{f_N(w_t|0,\Lambda_t^{-1})\ G_0(d\Lambda_t)}{g(w_t)}, && (13)\\
&\propto |\Lambda_t|^{1/2}\exp\left(-\frac{1}{2}w_t'\Lambda_t w_t\right)|\Lambda_t|^{(v+k-1-1-k)/2}\exp\left(-\frac{1}{2}tr(P^{-1}\Lambda_t)\right) && (14)\\
&\propto |\Lambda_t|^{(v-1)/2}\exp\left(-\frac{1}{2}tr([P^{-1}+w_tw_t']\Lambda_t)\right) && (15)\\
&\sim \text{Wishart}_k((w_tw_t'+P^{-1})^{-1}, v+k), && (16)
\end{aligned}
$$

and

$$
\begin{aligned}
g(w_t) &\equiv \int f_N(w_t|0,\Lambda_t^{-1})\ G_0(d\Lambda_t)\ d\Lambda_t && (17)\\
&\propto |w_tw_t'+P^{-1}|^{-(v+k)/2} && (18)\\
&\propto |P^{-1}|^{-(v+k)/2}|I_k+Pw_tw_t'|^{-(v+k)/2} && (19)\\
&\propto (1+w_t'Pw_t)^{-(v+k)/2} && (20)\\
&\propto (v+w_t'(Pv)w_t)^{-(v+k)/2} && (21)\\
&\sim St(w_t|0,(Pv)^{-1},v). && (22)
\end{aligned}
$$

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

$$
\begin{aligned}
B_j|\{w_t : s_t = j\} &\propto \left[\prod_{t:s_t=j}|B_t|^{1/2}\exp\left(-\frac{1}{2}w_t'B_tw_t\right)\right]|B_t|^{(v-2)/2}\exp\left(-\frac{1}{2}tr(P^{-1}B_t)\right)\\
&\propto |B_t|^{(v+n_j-2)/2}\exp\left(-\frac{1}{2}tr\left(\left[\sum_{t:s_t=j}w_tw_t'+P^{-1}\right]B_t\right)\right) && (23)\\
&\propto \text{Wishart}_k\left(\left(\sum_{t:s_t=j}w_tw_t'+P^{-1}\right)^{-1}, v+k+n_j-1\right), && (24)
\end{aligned}
$$

where $n_j = \#\{t : s_t = j\}$. This completes the DPM model portion of the Gibbs sampling.

Finally, if the precision parameter α is assigned a Gamma prior Gibbs sampling following Escobar & West (1995) can be used for the conditional posterior distribution.

[1] If $A \sim \text{Wishart}_k(S_0, v_0)$ then it has density

$$
p(A) = 2^{-v_0k/2}\pi^{-k(k-1)/4}|S_0|^{-v_0/2}\left[\prod_{i=1}^{k}\Gamma\left(\frac{v_0-1-i}{2}\right)\right]^{-1}|A|^{(v_0-1-k)/2}\exp\left(-\frac{1}{2}tr(S_0^{-1}A)\right). \quad (12)
$$

Next we sample the MGARCH model parameters. Let $\Gamma = (\Gamma_0, \Gamma_1, \Gamma_2)$. To sample from $\Gamma|B, S$ note that

$$\mathrm{Cov}(y_t|H_t, \Lambda_t) = H_t^{1/2}\Lambda_t^{-1}(H_t^{1/2})', \tag{25}$$

so that the conditional posterior is

$$p(\Gamma|\Lambda, Y_T) \propto \prod_{t=1}^{T} |H_t|^{-1/2}|\Lambda_t|^{1/2} \exp\left(-\frac{1}{2}y_t'(H_t^{-1/2})'\Lambda_t H_t^{-1/2} y_t\right) p(\Gamma). \tag{26}$$

This is a nonstandard density and we resort to Metropolis-Hastings sampler. Given the current value Γ, the proposal Γ' is sampled from

$$h(\Gamma') \sim \begin{cases} N(\Gamma, V) & \text{w.p.} \quad p \\ N(\Gamma, 100V) & \text{w.p.} \quad 1-p, \end{cases} \tag{27}$$

and accepted with probability

$$\min\{p(\Gamma'|\Lambda, y)/p(\Gamma|\Lambda, y), 1\} \tag{28}$$

and otherwise rejected and Γ is selected as the draw. This is a multivariate random walk directed by a mixture of normals. The second normal in the 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 $\log(p(\Gamma|\Lambda, Y_T))$ evaluated at the posterior mode.

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

The predictive density for w_t given the parameters follows the Polya urn prediction rule,

$$p(w_t|W, \Gamma, B, S) = \frac{\alpha}{\alpha+T} St(w_t|0, (P\nu)^{-1}, \nu) + \sum_{j=1}^{m} \frac{n_j}{\alpha+T} f_k(w_t|0, B_j^{-1}). \tag{29}$$

Applying the change of variables $y_t = H_t^{1/2} w_t$, with Jacobian term $|H_t|^{-1/2}$, gives the predictive density for y_t as,

$$\begin{aligned} p(y_t|Y_{t-1}, \Gamma, B, S) &= \frac{\alpha}{\alpha+T} St(y_t|0, H_t^{1/2}(P\nu)^{-1}(H_t^{1/2})', \nu) & (30)\\ &\quad + \sum_{j=1}^{m} \frac{n_j}{\alpha+T} f_k(y_t|0, H_t^{1/2} B_j^{-1}(H_t^{1/2})'). & (31) \end{aligned}$$

From this the covariance is

$$\mathrm{Cov}(y_t|Y_{t-1},\Gamma,B,S) \;=\; \frac{\alpha}{\alpha+T}\frac{H_t^{1/2}P^{-1}(H_t^{1/2})'}{\nu-2}+\sum_{j=1}^{m}\frac{n_j}{\alpha+T}H_t^{1/2}B_j^{-1}(H_t^{1/2})'. \quad (32)$$

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

$$p(y_t|Y_{t-1}) \approx \frac{1}{N}\sum_{i=1}^{N} p(y_t|Y_{t-1},\Gamma^{(i)},B^{(i)},S^{(i)}) \quad (33)$$

where $\{\Gamma^{(i)},B^{(i)},S^{(i)}\}_{i=1}^N$ are the posterior draws.

The predictive density of a portfolio of assets can also be conveniently derived from these results. Given the weights ω on wealth with $\sum_{i=1}^k \omega_i = 1$, the portfolio return is $y_t^p = \omega' y_t$ and has a predictive density

$$\begin{aligned} p(y_t^p|Y_{t-1},\Gamma,B,S,\omega) \;&=\; \frac{\alpha}{\alpha+T}St(y_t^p|0,\omega' H_t^{1/2}(P\nu)^{-1}(H_t^{1/2})'\omega,\nu) \quad (34)\\ &\quad+\sum_{j=1}^{m}\frac{n_j}{\alpha+T}f_k(y_t^p|0,\omega' H_t^{1/2}B_j^{-1}(H_t^{1/2})'\omega). \quad (35)\end{aligned}$$

Each of the distributions in this mixture are univariate. Averaging over the posterior draws as in (33) gives $p(y_t^p|Y_{t-1},\omega)$.

4 Approximate DPM mixture model

The previous model presented does not permit the mean of a component to affect the mixture. To do this we must depart from the conjugate results and the Polya urn formation of the DPM mixture.

Consider the same DPM mixture model as in Section 3 but conditional on H and written in terms of w_t,

$$\begin{aligned} w_t|\mu_t,\Lambda_t &\sim N(\mu_t,\Lambda_t^{-1}) \quad (36)\\ \mu_t,\Lambda_t|G &\overset{iid}{\sim} G, \quad (37)\\ G|G_0,\alpha &\sim \mathrm{DP}(G_0,\alpha), \quad (38)\\ G_0(\mu_t,\Lambda_t) &\equiv N(\underline{\beta},\underline{V}) - \mathrm{Wishart}_k(P,v+k-1),\; v\geq 1. \quad (39)\end{aligned}$$

Following Ishwaran & Zarepour (2000) the Dirichlet process can be approximated as $\mathcal{P}_L(.) = \sum_{\ell=1}^L p_\ell \delta_{\theta_\ell}(.)$ where $\theta_i = (\mu_i, B_i)$, $\Theta = \{\theta_1,\ldots,\theta_L\}$ are the unique elements independently and identically distributed as G_0 and L is finite. Completing the model is

$p|\alpha \sim \text{Dir}(\alpha/L, \dots, \alpha/L)$, for $0 \le p_\ell \le 1$ with $\sum_{\ell=1}^{L} p_\ell = 1$, the weights assigned to each θ_i. $\mathcal{P}_L$ converges as $L \to \infty$ to the Dirichlet process and can be used to approximate integrable functions of the process.

Let $\mu = (\mu_1, \dots, \mu_L)$, $B = (B_1, \dots, B_L)$ and $p = (p_1, \dots, p_L)$. The hierarchical form of the approximate mixture model is,

$$
\begin{aligned}
w_t|\Theta, S, \mu, B &\sim N(\mu_{s_t}, B_{s_t}^{-1}) && (40)\\
s_t|p &\sim \sum_{\ell=1}^{L} p_\ell \delta_\ell(.), && (41)\\
p|\alpha &\sim \text{Dir}(\alpha/L, \dots, \alpha/L), && (42)\\
\mu_{s_t} &\sim N(\underline{\beta}, \underline{V}) && (43)\\
B_{s_t} &\sim \text{Wishart}_k(P, v+k-1),\ v \ge 1, && (44)\\
\alpha &\sim p(\alpha). && (45)
\end{aligned}
$$

This leads to the following sequence of posterior draws,

- $S|p, \Theta, w$
- $p|\alpha, S$
- $\mu|B, S, w$
- $B|\mu, S, w$
- $\alpha|p$

A brief outline of the conditional distributions listed above are now given. For S we sequentially draw from the multinomial distribution with event probabilities

$$p(s_t = \ell|p, \Theta, w) \propto p(w_t|\theta_\ell)p_\ell, \quad \ell = 1, \dots, L, \tag{46}$$

for each t. Given the Dirichlet prior on p the conditional posterior is conjugate and

$$p(p|\alpha, S) \sim \text{Dir}(\alpha/L + n_1, \dots, \alpha/L + n_L), \tag{47}$$

where n_j was defined previously. Note that some n_j can be 0, as no observations are assigned to this cluster. The conditional posterior for μ is

$$p(\mu|B, S, w) \propto \prod_{\ell=1}^{L} \left[\prod_{t:s_t=\ell} p(w_t|\mu_\ell, B_\ell) \right] p(\mu_\ell). \tag{48}$$

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

$$\begin{aligned}
\mu_\ell|B,S,w &\sim N(\overline{\beta},\overline{V}) &(49)\\
\overline{\beta} &= \overline{V}\left(\underline{V}^{-1}\underline{\beta}+B_\ell\sum_{t:s_t=\ell}w_t\right) &(50)\\
\overline{V} &= \left(\underline{V}^{-1}+n_\ell B_\ell\right)^{-1} &(51)
\end{aligned}$$

For the conditional posterior of $B=\{B_1,\ldots,B_L\}$ we have[2]

$$p(B|\mu,S,w)\propto\prod_{\ell=1}^{L}\left[\prod_{t:s_t=\ell}p(w_t|\mu_\ell,B_\ell)\right]p(B_\ell). \tag{52}$$

Conjugacy makes this a series of Gibbs draws as in (24) but with w_t replaced by $(w_t-\mu_{s_t})$. For the empty sets $\{t:s_t=\ell\}$ none of the observations are allocated to this cluster and a direct draw from the Normal-Wishart (39) prior for μ_ℓ and B_ℓ is taken.

The last conditional distribution is nonconjugate but can be sampled by a Metropolis-Hastings step using

$$p(\alpha|p)\propto\frac{\Gamma(\alpha)}{\Gamma(\alpha/L)^L}p_1^{\alpha/L-1}\ldots p_L^{\alpha/L-1}p(\alpha). \tag{53}$$

This sequence of draws provides a valid sampler to the posterior distribution of the approximate DPM mixture model. Collecting the posterior draws in $\{S^{(i)},\Gamma^{(i)},\mu^{(i)},B^{(i)},p^{(i)}\}_{i=1}^N$ can be used for posterior inference and to compute the predictive density.

4.1 Predictive density

The structure in (40)-(45) implies the following predictive density conditional on model parameters,

$$p(y_t|Y_{t-1},\Gamma,\mu,B,p)=\sum_{\ell=1}^{L}p_\ell f_k(y_t|H_t^{1/2}\mu_\ell,H_t^{1/2}B_\ell^{-1}(H_t^{1/2})'). \tag{54}$$

Note that conditional on Y_{t-1} and model parameters, H_t is known. The predictive density with all parameter and density uncertainty integrated out can be estimated as

$$p(y_t|Y_{t-1})\approx\frac{1}{N}\sum_{i=1}^{N}p(y_t|Y_{t-1},\Gamma^{(i)},\mu^{(i)},B^{(i)},p^{(i)}) \tag{55}$$

where $\{\Gamma^{(i)},\mu^{(i)},B^{(i)},p^{(i)}\}_{i=1}^N$ are the posterior draws.

[2] This is not exactly the same definition of B used above as it includes redundant clusters that have no observations assigned to it.

5 Data

Two datsets, one of equity and the other foreign exchange are used to estimate the models. Data for equity returns is on IBM, the value-weighted CRSP 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.

6 Results

6.1 Estimates

For priors, each element of $\Gamma_0^{1/2}$, γ_1 and γ_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 model, $\alpha \sim Gamma(2, 8)$, $P = \frac{1}{\nu+k-1} I_k$ and $\nu = 10$. This implies $E[\Lambda_t] = I_k$. 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 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, is the MGARCH with a DPM model return distribution which allow the covariance of the components to differ. The last model approximates the stick breaking formulation of the DPM model and allows mixing over both mean and covariance of each mixture component. To implement this model we truncate the Dirichlet approximation of the DP at $L = 25$.

The MGARCH parameter are all very similar across the different models. Clearly, the 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 model features thick tails with the degree of freedom parameter 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 7 components, on average, in the DPM model. However, there is some posterior uncertainty as to the number of components with a density interval of $(5, 16)$ for the MGARCH-DPM model. The MGARCH-SB model estimates a smaller number of components. This is not surprising given that this model allows both the mean and covariance of the normal density to differ over components while the MGARCH-DPM only allows the covariances. In other words, the MGARCH-SB model is more flexible.

Figure 3 and 4 plot the conditional variances and the conditional correlations for the MGARCH-t and the MGARCH-DPM model. These quantities are derived from the posterior mean of the conditional covariances and for the DPM version are computed from (32) for a particular parameter draw. Both models have very similar patterns in their conditional moments through time.

6.2 Out-of-Sample Forecasts

In this section the forecast precision of the models is compared by predictive likelihoods. The results for equity, an equally weighted portfolio and for the FX data are reported in Table 3. This table does not report results for the MGARCH-SB model as we found them to be the same as the MGARCH-DPM model.[3] This indicates that asymmetry is not important to predicting this data.

For equity the semiparametric model offers a substantial improvement over the parameter model in prediction. The log-predictive Bayes factor in favour of the MGARCH-DPM is 14.64. The evidence for the MGARCH-DPM is weaker in the case of the portfolio with a log-predictive Bayes factor of 2.2. For the FX returns there is really no forecast gain compared to the MGARCH-t model as the semiparametric model matches the parametric model.

Why does the MGARCH-DPM model perform well? Some explanation of this can be found in the next two figures. Figure 5 reports several features of model comparison using the equity data. The top panel shows the period-by-period difference in the log-predictive likelihood. A positive (negative) value is in favor of the MGARCH-DPM (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.

[3] For example, the cumulative log predictive likelihood for the MGARCH-SB model is -1453.26 for equity data.

From this figure we can see 3 influential observations that are in favor of the MGARCH-DPM. They appear to be high volatility episodes. On the other hand, the second panel shows there to be regular ongoing forecast improvements from the MGARCH-DPM 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 6 displays the sequential estimates of the posterior mean of m, the number of clusters in the DPM mixture, along with average volatility. During the high volatility of the financial crisis the MGARCH-DPM 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 for a finite mixture model as the number of components is fixed.

7 Conclusion

This paper proposes a Bayesian nonparametric modeling approach for the return innovations in 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. Density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach.

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 Estadstica y Econometra.

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.

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), 503520.

Diamantopoulos, K. & Vrontos, I. D. (2010), 'A student-t full factor multivariate garch model', *Computational Economics* **35**(1), 63–83.

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.

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.

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.

Ishwaran, H. & Zarepour, M. (2002), 'Exact and approximate sum representations for the dirichlet process', *Canadian Journal of Statistics* **30**(2), 269283,.

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

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. & Ullah, A. (2005), Nonparametric and semiparametric multivariate garch model. University of California, Riverside.

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.

Sethuraman, J. (1994), 'A constructive definition of Dirichlet priors', *Statistica Sinica* **4**, 639–650.

Silvennoinen, A. & Tersvirta, T. (2007), Multivariate garch models, Working Paper Series in Economics and Finance 669, Stockholm School of Economics.
URL: *http://ideas.repec.org/p/hhs/hastef/0669.html*

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.

Table 1: Summary Statistics

	mean	var	skewness	kurtosis	Sample Correlation		
					IBM	VW	HP
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
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

Table 2: Full Sample Estimates, CRSP Data

	MGARCH-t		MGARCH-DPM		MGARCH-SB	
g_{11}	0.1017	(0.0748, 0.1288)	0.1200	(0.0849, 0.1545)	0.1247	(0.0879, 0.1619)
g_{21}	0.0581	(0.0362, 0.0856)	0.0317	(-0.0152,0.0727)	0.0451	(0.0101, 0.0850)
g_{31}	0.0418	(0.0025, 0.0816)	0.0084	(-0.0741, 0.0988)	0.0268	(-0.0598, 0.1073)
g_{22}	0.0623	(0.0453, 0.0795)	0.0732	(0.0533, 0.0933)	0.0792	(0.0563, 0.1039)
g_{32}	0.0926	(0.0518, 0.1357)	0.0996	(0.0377, 0.1723)	0.0999	(-9.0e-5, 0.1750)
g_{33}	0.1633	(0.1075, 0.2147)	0.1214	(0.0428, 0.1790)	0.1261	(2.4e-3, 0.1961)
γ_{11}	0.1534	(0.1341, 0.1735)	0.1914	(0.1626, 0.2198)	0.1932	(0.1593, 0.2324)
γ_{12}	0.1883	(0.1619, 0.2165)	0.1936	(0.1615, 0.2228)	0.2056	(0.1723, 0.2392)
γ_{13}	0.1636	(0.1428, 0.1865)	0.1570	(0.1318, 0.1839)	0.1664	(0.1385, 0.2000)
γ_{21}	0.9793	(0.9737, 0.9838)	0.9777	(0.9721, 0.9823)	0.9780	(0.9719, 0.9828)
γ_{22}	0.9721	(0.9620, 0.9796)	0.9649	(0.9577, 0.9719)	0.9662	(0.9575, 0.9738)
γ_{23}	0.9787	(0.9731, 0.9836)	0.9746	(0.9687, 0.9796)	0.9760	(0.9702, 0.9810)
ψ	7.7147	(6.5995, 8.9348)				
α			0.6784	(0.2391, 1.3544)	0.6241	(0.1758, 1.2689)
m			9.4901	(5.0000, 16.000)	7.4888	(3.0000, 12.0000)

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

Table 3: Cumulative Log-Predictive Likelihoods

	Equity	Portfolio	FX
MGARCH-t	-1468.26	-548.39	-1096.47
MGARCH-DPM	-1453.62	-546.19	-1096.63
Observations	262	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.

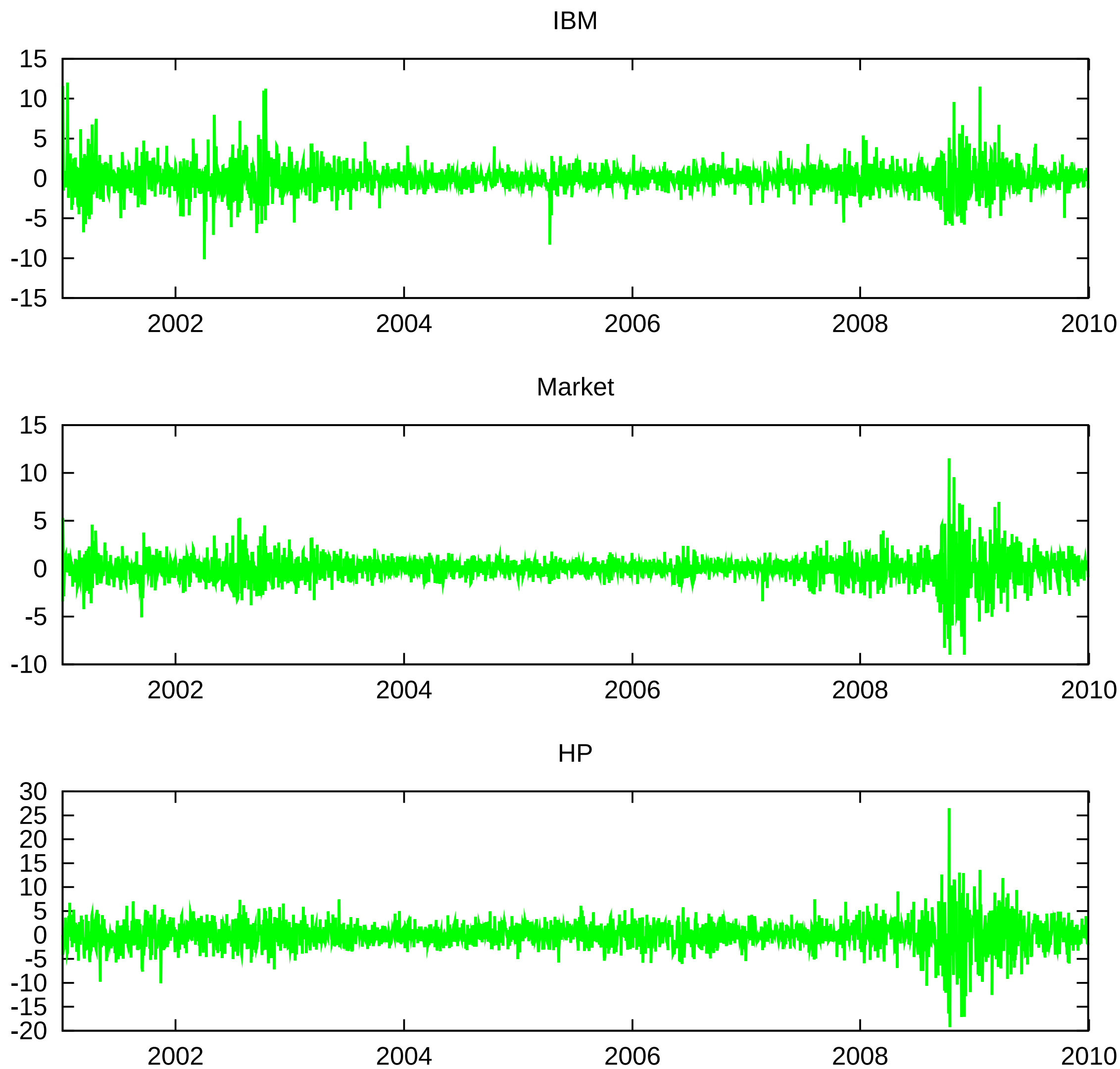

Figure 1: Daily Equity Returns

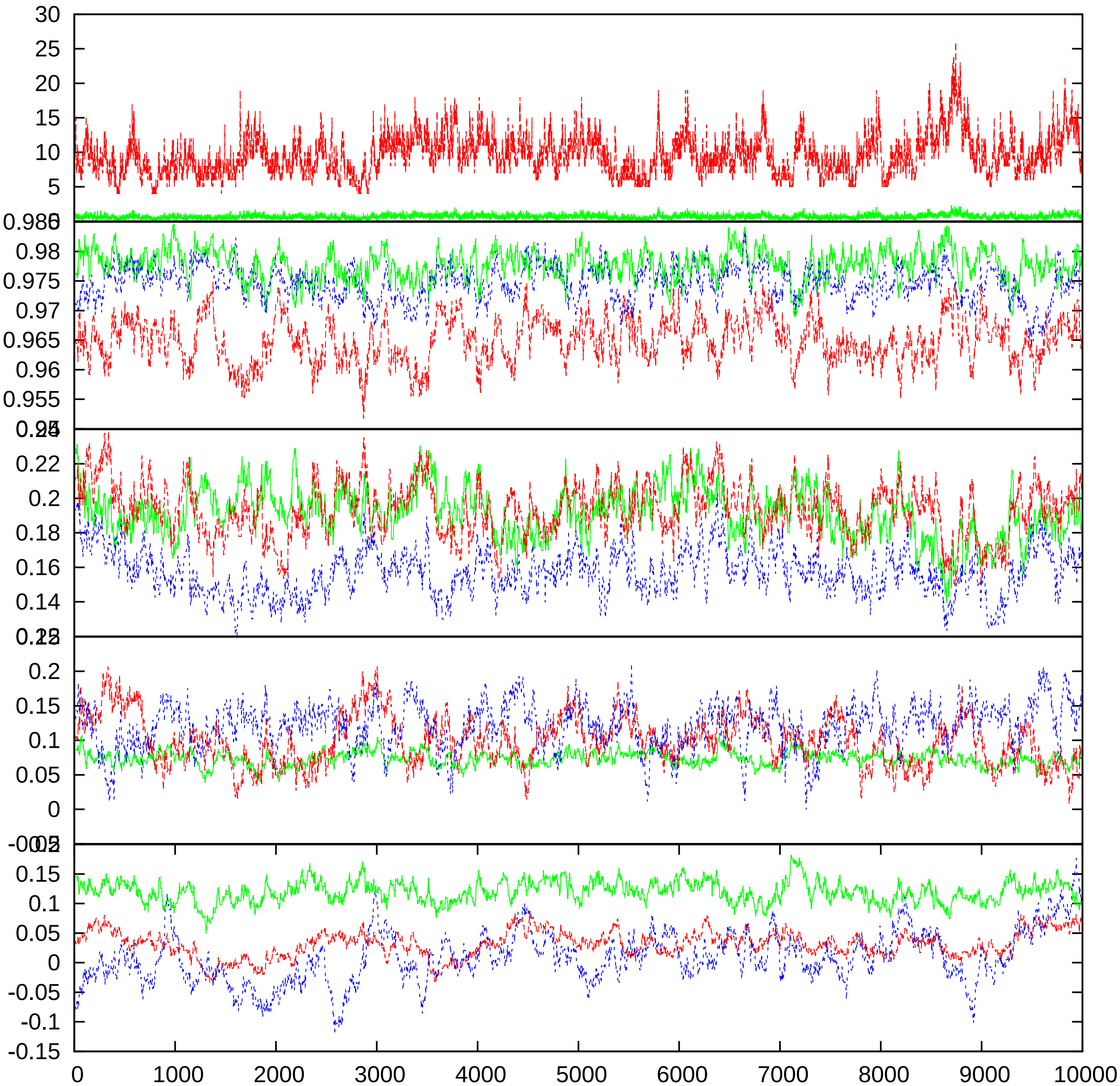

Figure 2: Trace plots of parameters

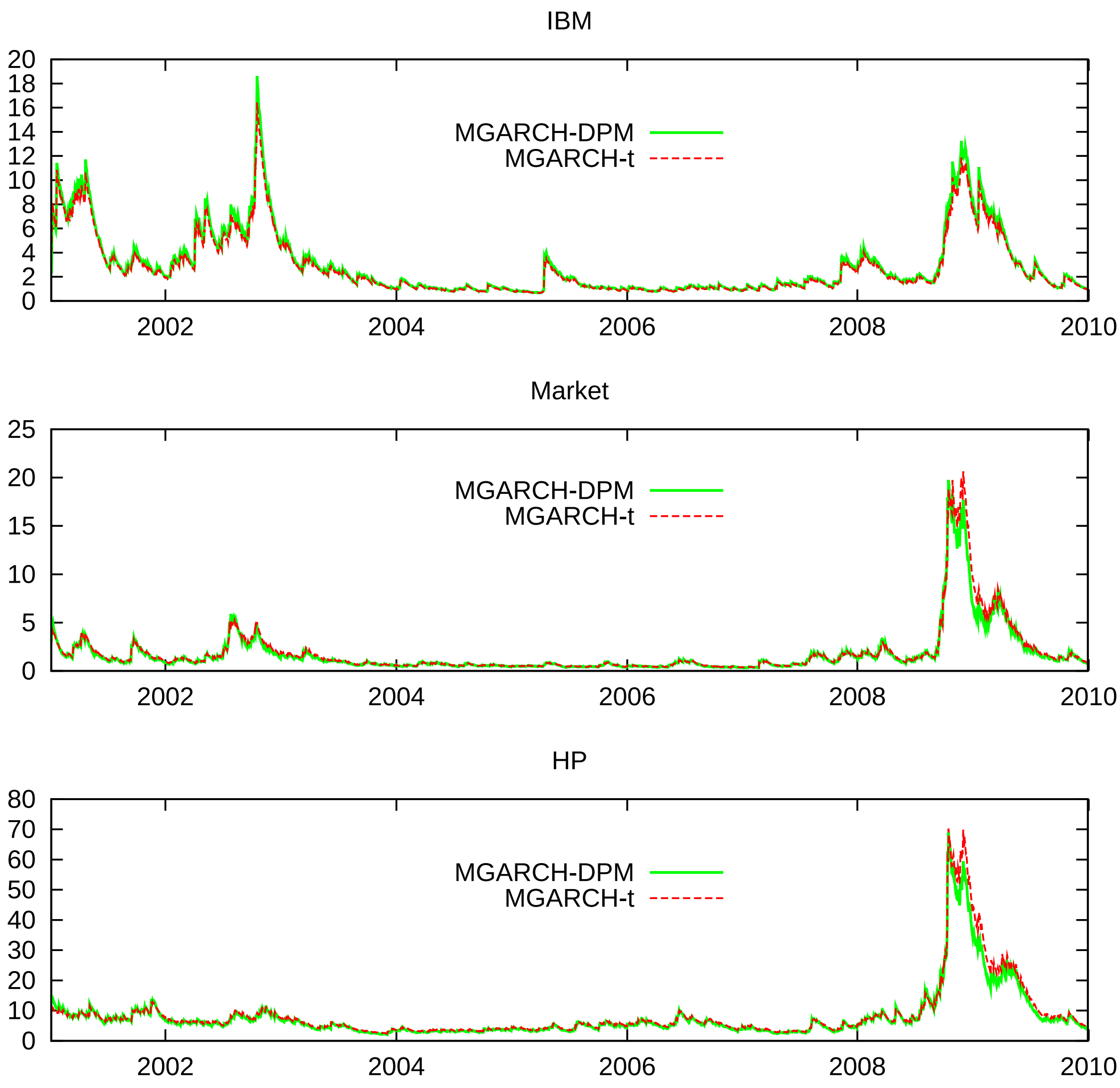

Figure 3: Conditional Variances

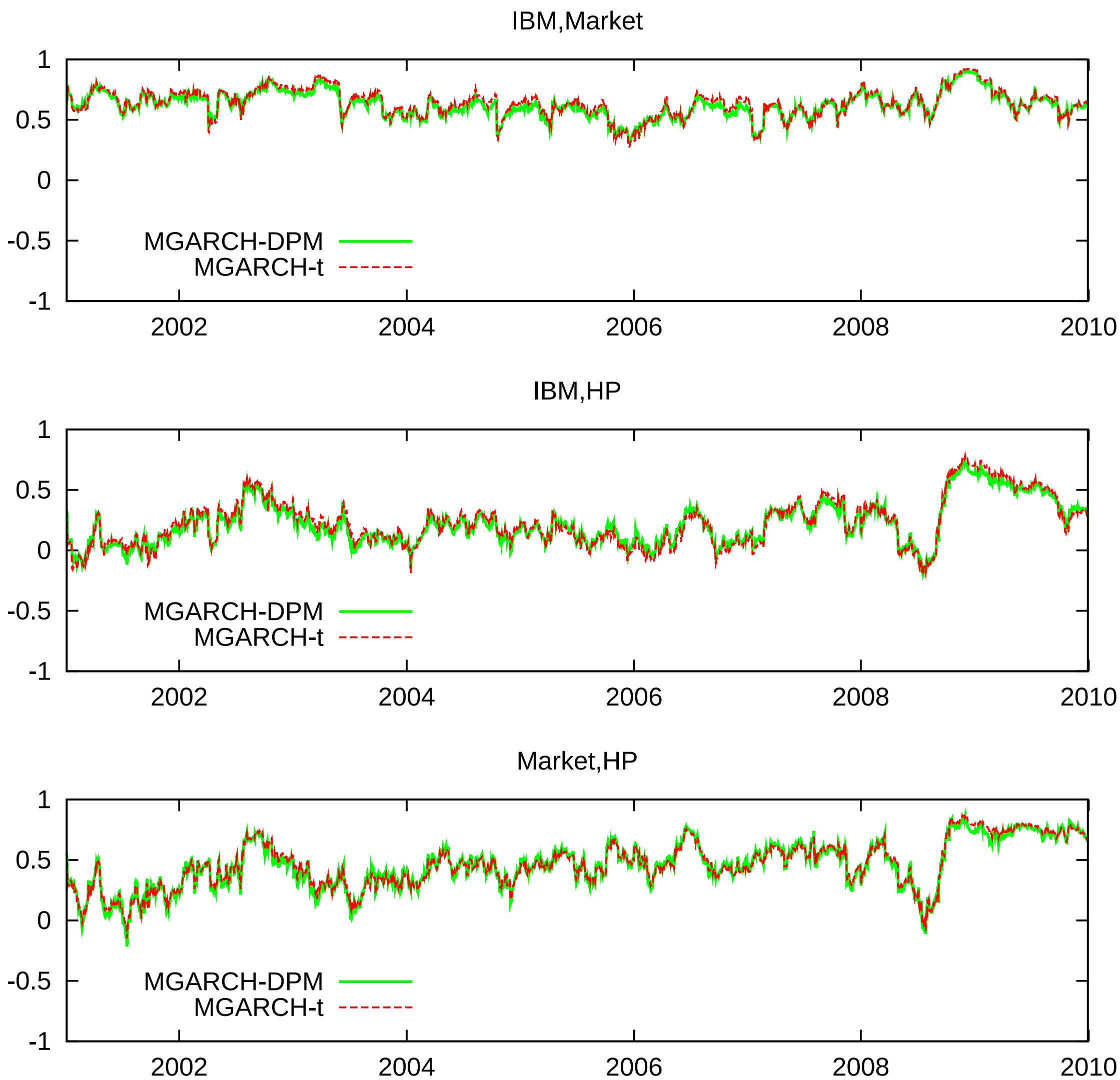

Figure 4: Conditional Correlations

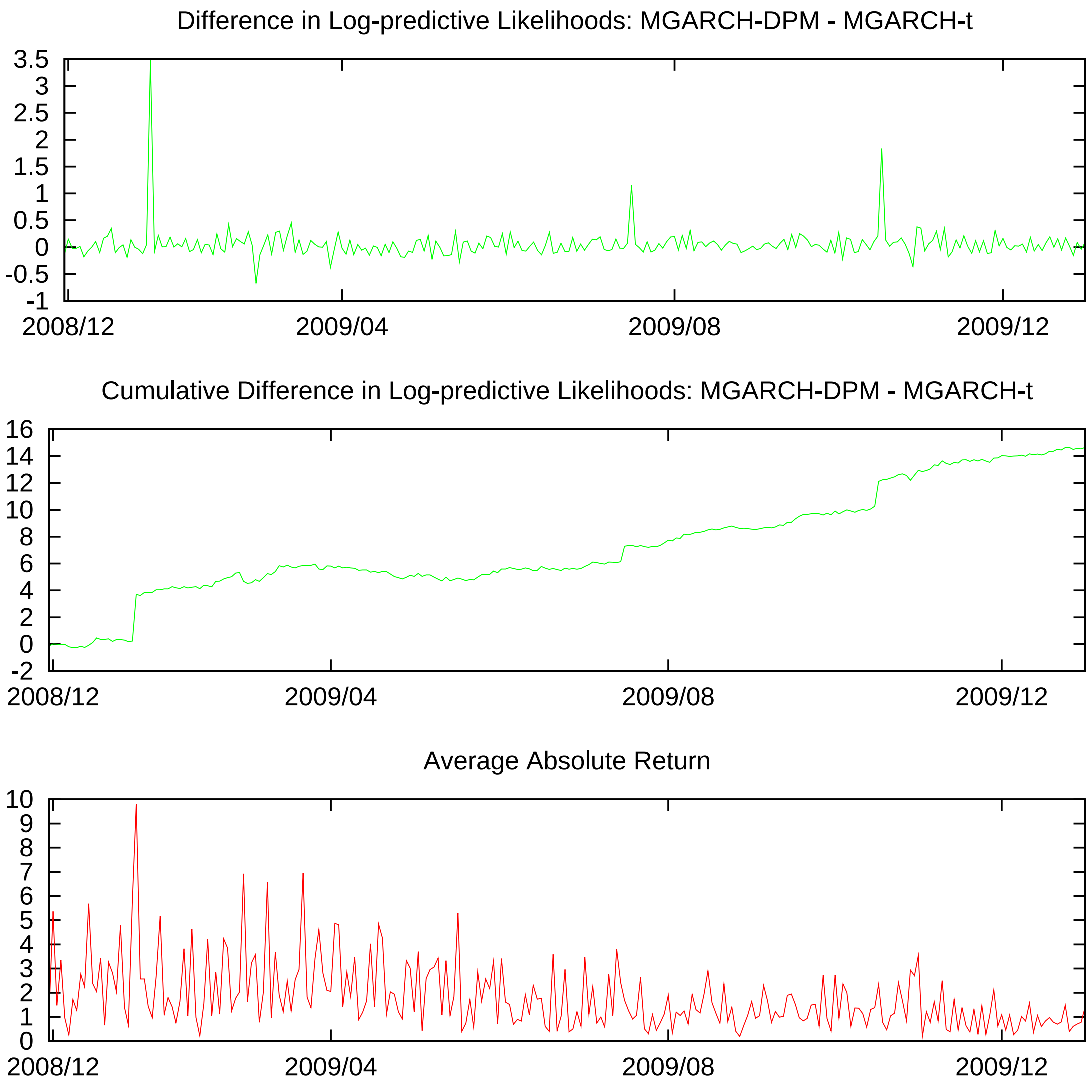

Figure 5: Model Comparison

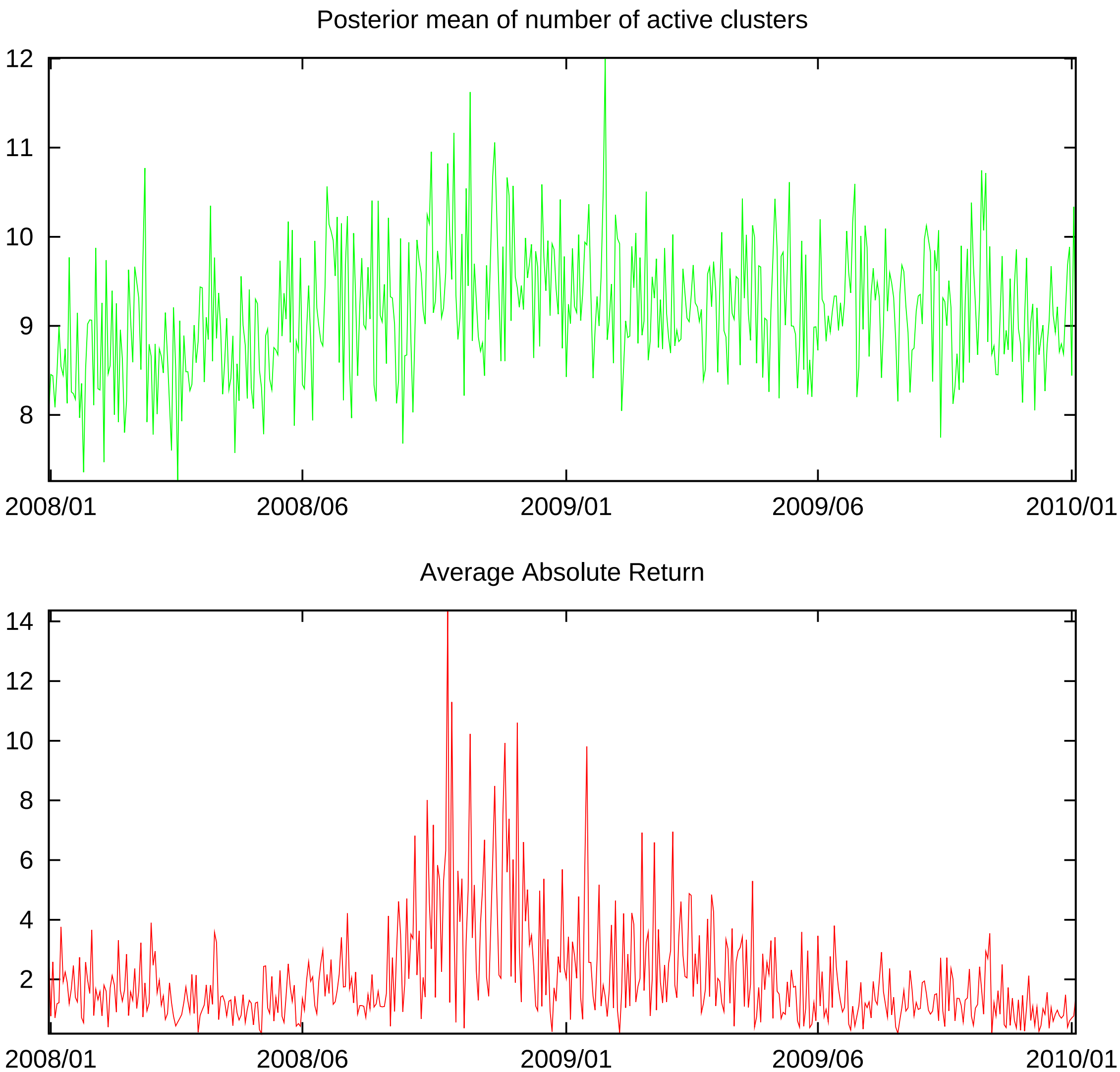

Figure 6: Sequential Estimates 2008-2009/12