Bayesian Markov mixture of normals approach to modeling financial returns

George Chang
Department of Finance and Quantitative Methods, Foster College of Business, Bradley University, Peoria, Illinois, USA

Abstract

Purpose – The purpose of this paper is to investigate whether Markov mixture of normals (MMN) model is a viable approach to modeling financial returns.

Design/methodology/approach – This paper adopts the full Bayesian estimation approach based on the method of Gibbs sampling, and the latent state variables simulation algorithm developed by Chib.

Findings – Using data from the S&P 500 index, the paper first demonstrates that the MMN model is able to capture the unconditional features of the S&P 500 daily returns. It further conducts formal model comparisons to examine the performance of the Markov mixture structures relative to two well-known alternatives, the GARCH and the t-GARCH models. The results clearly indicate that MMN models are viable alternatives to modeling financial returns.

Research limitations/implications – The univariate MMN structure in this paper can be generalized to a multivariate setting, which can provide a flexible yet practical approach to modeling multiple time series of assets returns.

Practical implications – Given the encouraging empirical performance of the MMN models, it is hopeful that the MMN models will have success in some interesting financial applications such as Value-at-Risk and option pricing.

Originality/value – The paper explicitly formulates the Gibbs sampling procedures for estimating MMN models in a Bayesian framework. It also shows empirically that MMN models are able to capture the stylized features of financial returns. The MMN models and their estimation method in this paper can be applied to other financial data, especially in which tail probability is of major interest or concern.

Keywords Markov processes, Bayesian statistical decision theory

Paper type Research paper

Financial asset returns often exhibit time-varying conditional distributions, departures from normality, and persistence in volatility. Various empirical studies of financial asset returns have evidenced the temporal properties that daily returns are not autocorrelated; albeit, daily absolute returns and daily squared returns exhibit slowly decaying autocorrelations. Higher moments of financial returns distribution are especially important in applications in which the tail probability is of major interest or concern. Markov mixture of distributions allows for both temporal dependence and stochastic variation in the moments. Markov mixtures models are capable of handling nearly any

The author is grateful to John Geweke, David Bates, Mike Stutzer, Paul Weller, and Garland Durham for helpful comment and discussions. He has also benefited from comments of participants at the FMA meetings, finance seminar, and applied statistics seminar at The University of Iowa. The usual disclaimers apply.

Studies in Economics and Finance
Vol. 23 No. 2, 2006
pp. 141-158

1086-7376
DOI 10.1108/10867370610683923

feature of the unconditional distribution including leptokurtosis and skewness. Moreover, mixtures of normals are very tractable since the normal theory is well known.

Independent mixture models and even Markov mixtures are not recent innovations. The present form of the Markov mixture structure is given in Lindgren (1978). In spite of their analytical appeal, the application of Markov mixtures in econometrics and empirical finance has been limited. To a larger extent, the limited applications can be explained by the fact that, in any sample, each observation generates a singularity in the likelihood function. However, this problem can be addressed using Bayesian Markov chain Monte Carlo (MCMC) methods.

This paper explicitly formulates the Gibbs sampling procedures for estimating Markov mixture of normals (MMN) models. The estimation and inference of the MMN structures adopts the Bayesian Markov chain simulation method. In particular, the Gibbs sampling-data augmentation (GS-DA) algorithm by Chib (1996) is employed. We first show that MMN model is able to capture the unconditional features of the S&P 500 index daily returns. Given this encouraging performance of MMN models to modeling financial returns, it is hopeful that the MMN models will have success in other interesting financial applications such as Value-at-Risk (VaR), and option pricing. Further, we apply the S&P 500 index data to conduct formal model comparison to address the issues of model specification and the performance of the Markov mixture structures relative to certain well known alternatives. Using the same data, marginal likelihood approximations for the GARCH(1,1) model with Gaussian innovations (referred to as GARCH), the GARCH(1,1) model with Student-*t* innovations (t-GARCH), the 2-state MMN model (MMN(2)), the 3-state MMN model (MMN(3)), and the 4-state MMN model (MMN(4)) are obtained.

The rest of the paper proceeds as follows. Section 1 presents the effectiveness of the Gibbs sampling estimation procedures for MMN models. Section 2 describes some alternative discrete-time data generating processes. Section 3 provides a description of the S&P 500 index data. Section 4 conducts prior calibrations for model comparison purposes. Section 5 performs formal model comparison analysis. Section 6 offers concluding remarks.

1. MMN for asset return dynamics

The univariate MMN model can be described as follows:

$$y_t|(\theta, s_t = i) \sim N\left(\mu_i, h_i^{-1}\right) \quad i = 1, \ldots, m \tag{1}$$

$$s_t|s_{t-1} \sim \text{Markov}(P, \pi_1)$$

where s_t is the state indicator, θ is the set of parameters in the model, $P = \{p_{ij}\}$ is the one-step transition probability matrix of the chain, i.e. $p_{ij} = Pr(s_t = |s_{t-1} = i)$, and π_1 is the probability distribution at $t = 1$.

Let $Y_T = \{y_1, \ldots, y_t\}$, $\phi_i = \{\mu_i, h_i\}$, and $\theta = \{\phi_1, \ldots, \phi_m, P, \pi_1\}$, then the normalized density is given by:

$$f(Y_T|\theta) = \prod_{t=1}^{T} f(y_t|Y_{T-1}, \theta) \tag{2}$$

where

$$f(y_t|Y_{T-1},\theta) = \sum_{j=1}^{m} f(y_t|Y_{T-1},\phi_j)p(s_t = j|Y_{t-1},\theta)$$

A major problem with estimating Markov mixture models is that the likelihood function is known to be very ill-conditioned, and possibly encountering unbounded likelihood. Different approaches have been proposed for parameter estimation. A quite different approach to the estimation of mixture models is possible from the Bayesian Markov chain simulation perspective. The Bayesian Markov chain simulation method is free of the aforementioned drawbacks of the maximum likelihood estimation method. The Bayesian methods are applicable to the class of finite Markov mixtures of distributions that admit conditionally conjugate structures, including mixtures of normal distributions. The basic idea is that the computation of the likelihood function can be avoided if the population index variable $\{s_t\}$ is treated as an unknown parameter and simulated alongside the other parameters of the model by Gibbs sampling methods. Albert and Chib (1993) and McCulloch and Tsay exploit this idea to simulate the posterior distribution. Chib (1996) provides a GS-DA algorithm supporting inference for Markov mixtures of normals.

The estimation approach adopted in this paper is the full Bayesian approach based on the method of Gibbs sampling, and the latent state variables simulation algorithm developed by Chib (1996). Derivations of the Gibbs sampling procedures used in this paper are provided in the Appendices.

To examine the effectiveness of the estimation method to recovering the true distribution, simulated data from some known MMN processes are used[1]. Since, parameters in this model are not invariant to label switching, the emphasis here is to examine the derived predictive distributions which are invariant to this problem. That is, we examine whether or not the overall (unconditional) statistics estimated from the MMN model are close to their population counterparts[2]. Artificial samples of size 1,000, are generated from the following Markov mixture of two normals (MMN(2)) process, where s_t is the state indicator at time t, P is the transition probability matrix, and π_1 is the stationary probability distribution:

$$y_t|s_{t=1} \sim N(0,9) \quad y_t|s_{t=2} \sim N(0,1) \quad P = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} \quad \pi_1 = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}' \tag{3}$$

As seen from Table I, this estimation method is able to recover the true distibution subject to the random sampling error of the artificial data. Artificial data sets from a number of MMN processes with different means, and standard deviations have also been examined. The effectiveness of the estimation method is confirmed. It should be noted that as the mixture components distributions become less distinct from one another, more observations are needed in order to get fairly accurate estimates, which is inevitable regardless of what estimation technique is used.

2. Some alternative returns distributions

Important innovations addressing the conditional features of asset returns include the ARCH model of Engle (1982) and the GARCH model given in Bollerslev (1986), which provide parametric expressions for the temporal dependencies in the variances. GARCH(1,1) model has often been found to appropriately account for conditional heteroskedasticity in the empirical analysis of financial data. Numerous extensions to

Table I. Unconditional statistics estimated directly from the sample vs unconditional statistics estimated from MMN(2) model

	Mean	Variance	Skewness	Kurtosis	ρ_1^{a}	Max	Min
D[b]	−0.1003	4.8459	−0.1261	4.2587	0.0027	8.0708	−7.9327
M[c]	−0.1017	4.8617	−0.0479	4.4148	0.0008		
	(0.0687)	(0.3198)	(0.0723)	(0.2360)	(0.0011)		
	$\hat{P}$		$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\pi}$
D	$\begin{bmatrix} 0.7881 & 0.2119 \\ 0.2827 & 0.7173 \end{bmatrix}$		−0.1504	−0.0337	7.7931	0.9244	$\begin{bmatrix} 0.5716 \\ 0.4284 \end{bmatrix}$
M	$\begin{bmatrix} \underset{(0.0498)}{0.7843} & \underset{(0.0498)}{0.2157} \\ \underset{(0.0623)}{0.3029} & \underset{(0.0623)}{0.6971} \end{bmatrix}$		−0.1451	−0.0404	7.6851	0.9022	$\begin{bmatrix} \underset{(0.0512)}{0.5846} \\ \underset{(0.0512)}{0.4154} \end{bmatrix}$
			(0.1212)	(0.0707)	(0.6324)	(0.1575)	

Notes: [a] ρ_1 denotes first-order autocorrelation coefficient; [b] D denotes estimates computed directly from the simulated data, i.e. they are simply the sample analogs of the population statistics; [c] M denotes estimates obtained from fitting MMN(2) model to the simulated data. The reported estimates from the MMN(2) model are the posterior means, and standard deviations (in parentheses). Diffuse priors are employed. Calculations of the unconditional statistics for MMN models are given in Appendix 2 these models have appeared in the literature. Examples include the GARCH model with Student-t innovations (Bollerslev, 1986), the ARCH-M model (Engle and Bollerslev, 1986), and the EGARCH model (Nelson). An inherent limitation of the ARCH and GARCH models arises from the fact that the evolution of the conditional variance is deterministic, and thus, sudden and extreme changes are difficult to accommodate. Another undesirable feature of the GARCH models is that the long-term (many-period ahead) predictive distributions from those models have implications that are unreasonable, as discussed in Geweke and McCausland (2001). Other more recent times series models for financial returns include the stochastic volatility model by Jacquier *et al.* (1994), where the log of the conditional variance follows an AR process with Gaussian innovations.

The GARCH model introduced by Bollerslev (1986) is a leading model that captures some of the stylized facts for a single asset return. Denote the observable return from period $t-1$ to period t by y_t, and the variance of y_t at time t by h_t. The GARCH(1,1) model has the process:

$$y_t = \mu + \varepsilon_t \tag{4}$$

$$\varepsilon_t = \sqrt{h_t} \cdot v_t \tag{5}$$

$$h_t = \alpha + \delta\varepsilon_{t-1}^2 + \gamma h_{t-1} \tag{6}$$

$$v_t \overset{IID}{\sim} N(0,1)$$

$$\alpha > 0, \delta \geq 0, \gamma \geq 0, \delta + \gamma < 1.$$

The GARCH model with Gaussian innovations tends to display an unconditional distribution with relatively lower weight in the tails compared to the observed data distribution. This insufficiency in the tail weight can be improved by replacing the Gaussian innovations with Student-t innovations. Hence, the t-GARCH model:

$$y_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t} \cdot v_t$$

$$h_t = \alpha + \delta\varepsilon_{t-1}^2 + \gamma h_{t-1}$$

$$v_t \overset{IID}{\sim} t(0, s, \nu) \tag{7}$$

$$\alpha > 0, \delta \geq 0, \gamma \geq 0, \delta + \gamma < 1.$$

where s is the scale parameter for Student-t distribution. Note that as the degrees of freedom parameter ν decreases, the tails of the distribution fatten. Also the conditional moments of y_t exist only up to order ν. The t-GARCH model is widely regarded as the best fitting simple model for financial asset returns.

3. Data description

Daily returns of the S&P500 from January 2, 1981 to December 31, 1991 are used for this study. The sample consists of 2,781 daily (log) return observations. Subseries of the sample are examined. Series-I consists of daily returns from 1981 to 1986. Series-II consists of daily returns from 1987 to 1991. Series-III consists of the entire sample from 1981 to 1991. In order to get an idea of how much impact the extreme daily return of October 1987 market crash has on the unconditional statistics in Series-II and Series-III, two other series are examined. Series-IV and Series-V are constructed from Series-II and Series-III, respectively, but with the October 1987 market crash excluded. Summary statistics for these five series are provided in Table II. The hypothesis that the returns in each series are drawn from a normal distribution is strongly rejected by Anderson-Darling normality test.

Some salient features of these S&P 500 daily returns series include:

- The average daily returns are almost zero for all series, and are negligible compared to the standard deviations of the daily returns. In other words, the daily returns are quite noisy.
- The daily returns have asymmetric distributions, with series I slightly skewed to the right, and Series II and III skewed to the left.
- The daily returns have fatter tails than those of the normal distribution. Comparing Series IV and V with Series II and III, we see that much of the skewness and kurtosis are driven by the stock market crash in October 1987.

To examine the ability of Markov mixture models to capture the unconditional distribution of the daily returns series, we fit various MMN models to Series I, II, and III, and compute the resulting unconditional statistics of the data series. The priors used in this analysis are those from priors calibrations detailed in the next section[3]. The results (posterior means of the estimated unconditional statistics) are presented in Table III. We see that within each data series, increasing the number of mixture components improves the overall fit in terms of the unconditional statistics. MMN(4) has the best overall fit for Series I, and MMN(5) has the best overall fit for Series II and III. The improvement of MMN(5) over MMN(4) is most obviously seen in Series II and III which contain the extreme value of the 1987 stock market crash. In fact, the estimated MMN(5) is characterized by a normal distribution with mean being about -0.23 and standard deviation being about 0.01 (Table IV.) This state essentially captures the 1987 stock market crash as a crash state, much in the same spirit as a negative jump in some continuous time models.

4. Priors

To minimize the impact of the priors on the model comparison analysis later, the parameters for the priors are selected such that the priors for each model reflect similar beliefs regarding the unconditional distribution of $\{y_t\}$ and the general level of persistence,

Series	No. of obs	Mean	SD	Skew	Kurt	Max	Min
I	1,517	0.00038	0.0089	0.1320	5.06	0.0465	-0.0493
II	1,264	0.00043	0.0128	-4.8497	89.54	0.0871	-0.2289
III	2,781	0.00040	0.0108	-3.5864	80.32	0.0871	-0.2289
IV	1,263	0.00061	0.0110	-0.4358	12.28	0.0871	-0.0864
V	2,780	0.00049	0.0099	-0.2136	10.32	0.0871	-0.0864

Note: Series I (1981-1986), II (1987-1991), and III (1981-1991); Series IV (1987-1991), and V (1981-1991), October 1987 crash excluded

Table II.
Summary statistics for S&P daily returns (1981-1991)

Table III. Posterior means of the statistics estimated from Markov mixture models

Series	Model	Mean	SD	Skew	Kurt	Max	Min
	Empirical	0.0004	0.0089	0.132	5.06	0.0465	− 0.0493
I	MMN(2)	0.0003	0.0089	0.012	4.13	–	–
(1981-1986)	MMN(3)	0.0003	0.0089	0.022	4.37	–	–
	MMN(4)	0.0003	0.0089	0.024	4.60	–	–
	Empirical	0.0004	0.0128	− 4.850	89.54	0.0871	− 0.2289
	MMN(2)	0.0006	0.0133	− 0.108	15.14	–	–
II	MMN(3)	0.0006	0.0148	− 0.103	40.18	–	–
(1987-1991)	MMN(4)	0.0006	0.0148	− 0.101	40.31	–	–
	MMN(5)	0.0003	0.0128	− 5.201	90.65	–	–
	Empirical	0.0004	0.0108	− 3.586	80.32	0.0871	− 0.2289
	MMN(2)	0.0004	0.0110	− 0.066	10.82	–	–
III	MMN(3)	0.0005	0.0119	− 0.064	40.10	–	–
(1981-1991)	MMN(4)	0.0005	0.0119	− 0.048	41.21	–	–
	MMN(5)	0.0003	0.0124	− 3.416	60.68	–	–

Table IV. MMN(5) parameter values with highest posterior density

(1 = 0.00036	N1 (− 0.2298, 0.0091)
(2 = 0.0021	N2 (0.0657, 0.0100)
(3 = 0.0036	N3 (0.0528, 0.0093)
(4 = 0.2565	N4 (0.0007, 0.0128)
(5 = 0.7374	N5 (0.0005, 0.0077)

$$P = \begin{bmatrix} 0.0058 & 0.8827 & 0.0170 & 0.0532 & 0.0412 \\ 0.0002 & 0.7631 & 0.1581 & 0.0228 & 0.0558 \\ 0.0281 & 0.0267 & 0.0639 & 0.8812 & 0.0001 \\ 0.0001 & 0.0001 & 0.0077 & 0.9505 & 0.0413 \\ 0.0003 & 0.0001 & 0.0014 & 0.0128 & 0.9854 \end{bmatrix}$$

Note: Estimation data period: 1981-1991 (2,781 daily returns)

and that each prior be sufficiently diffuse. It should be noted that improper (flat) priors should never be used in Bayesian model comparisons, due to Lindley's paradox, after Lindley and Bartlett[4]. Further, given the large data sample size used in this paper, the likelihood will dominate the prior in determining the posterior distribution.

In relation to the unconditional distribution, the priors for each model are chosen such that the unconditional mean, standard deviation, and kurtosis of $\{y_t\}$ (in terms of prior expectations) are roughly the same[5]. Specifically, we require the quantiles of the moments of interest to be roughly the same across models. For the MMN model, the persistence of volatility is accomplished by requiring that $E_{p(\theta)}(P)$ place most weight on the diagonal. To ensure that the intended features were adequately supported, it was verified that the expected autocorrelation functions for $\{|y_t|\}$ and $\{y_t^2\}$ did not die out too quickly. These objectives were tempered by requiring that $E_{p(\theta)}(\pi)$ not be dominated by any single state.

These calibrations and approximations were carried out by making i.i.d. draws of the entire parameter vector from the prior, then proceeded in the obvious manner.

The historical data used for calibrations consist of 1,000 daily returns, which roughly correspond to the four-year period prior to the data Series I. Table V presents the means and quantiles (in brackets) of the predictive distributions of interest for each model through prior calibrations.

5. Bayes factors for model comparisons

The determination of the optimal specification of Markov mixtures, i.e. the selection of the number of states and the specification of the set of mixing components can be conducted by Bayes factors[6]. The Bayes factor in favor of model i over model j based upon Y_T,denoted by $B_{i,j}(Y_T)$, is simply $(M_i(Y_T))/(M_j(Y_T))$, where $M_j(Y_T)$ is the marginal likelihood for model i. Specifically:

$$M_i(Y_T) = \int_{\Theta} p(\theta)p(Y_T|\theta)\,\mathrm{d}\theta \text{ where } \Theta \text{ is the prior parameter space} \tag{8}$$

The calculation of the marginal likelihood, which is the normalizing constant of the posterior density, was once challenging. However, several approaches have been developed in recent years that make the calculation of the marginal likelihood possible. This paper uses the approach developed by Chib (1995), which is applicable to our context of Bayes estimation via Gibbs sampling. This approach exploits the fact that the marginal density can be expressed as the prior times the likelihood function over the posterior density, and this identity holds for any parameter value:

$$M(Y_T) = \frac{p(Y_T|\theta)p(\theta)}{p(\theta|Y_T)} \quad \forall\theta \in \Theta \tag{9}$$

Consequently, the log of the marginal likelihood is[7]:

$$\ln \hat{M}(Y_T) = \ln p(Y_T|\theta^*) + \ln p(\theta^*) - \sum_{r=1}^{B} \ln \hat{p}\left(\theta_r^*|Y_T, \theta_s^*(s<r)\right) \tag{10}$$

where:

$$\hat{p}\left(\theta_r^*|Y_T, \theta_s^*(s<r)\right) = G^{-1}\sum_{j=1}^{G} p\left(\theta_r^*|Y_T, \theta_1^*, \theta_2^*, \ldots, \theta_{r-1}^*, \theta_l^{(j)}(l>r), z^{(j)}\right)$$

and $(\theta_1, \theta_2, \ldots, \theta_B)$ are the Gibbs sampling blocks, and z denotes the latent data.

	$E(y_t)$	$\sqrt{Var(y_t)}$	Kurt (y_t)	$\mathrm{Corr}(y_t^2, y_{t+1}^2)$
Old data	0.0000	0.0094	4.9561	0.1652
GARCH	0.0000[a]	0.0144	4.0805	0.0907
	[−0.0006 0.0006]	[0.0035 0.0158]	[3.0102 3.2878]	[0.0214 0.1286]
t-GARCH	0.0000	0.0160	7.5793	0.0887
	[−0.0007 0.0008]	[0.0039 0.0167]	[3.8311 5.6201]	[0.0180 0.1316]
MMN(2)	0.0000	0.0152	4.7010	0.0701
	[−0.0005 0.0006]	[0.0072 0.0160]	[3.1037 4.4956]	[0.0132 0.1124]
MMN(3)	0.0000	0.0159	4.8341	0.0844
	[−0.0004 0.0005]	[0.0099 0.0173]	[3.3547 5.133]	[0.0355 0.1239]
MMN(4)	0.0000	0.0159	5.3765	0.0944
	[−0.0004 0.0004]	[0.0105 0.0178]	[3.6446 5.9537]	[0.0549 0.1341]

Note: [a]Means and quantiles (in brackets) of the quantity of interest

Table V.
Predictive distributions of interest through prior calibrations

5.1 Results and discussions

The log-marginal likelihoods for the competing models are reported in Table VI. For all series, the GARCH log-marginal likelihoods are the lowest, and the magnitude of the differences is relatively large. For example, the Bayes factors in favor of MMN(4) over GARCH range from 3.56×10^{11} for Series I to 1.34×10^{42} for Series II. The t-GARCH model has the greatest log-marginal likelihood for Series I and III, while the MMN(4) has the greatest log-marginal likelihood for Series II. For Series I, the MMN(3) and MMN(4) models are nearly identical under the Bayes factor standard. However, for Series II and III, MMN(4) has significantly larger log-marginal likelihood than MMN(3). This better fit of the MMN(4) over MMN(3) is because the additonal state in the MMN(4) provides a good fit to the extreme values of the data; whereas in the MMN(3), the three states have to really stretch out to fit the extreme data values.

6. Concluding remarks

This paper explicitly formulates the Gibbs sampling procedures for estimating MMN models. The estimation and inference of the MMN structures adopts the Bayesian Markov chain simulation method. In particular, the GS-DA algorithm by Chib (1996) is employed. We first show that MMN model is able to capture the unconditional features of the S&P 500 index daily returns. Further, we apply the S&P 500 index data to conduct formal model comparison to address the issues of model specification and the performance of the Markov mixture structures relative to certain well known alternatives. Using the same data, marginal likelihood approximations for the GARCH, t-GARCH, MMN(2), MMN(3) and MMN(4) are obtained.

We found that the GARCH model has by far the lowest marginal likelihood. The other models are in a horse race in their marginal likelihood, with either t-GARCH or the leading MMN model having the highest marginal likelihood depending on the data series. Furthermore, the predictive log Bayes factors analysis suggests that t-GARCH and the leading MMN models perform, roughly speaking, equally well with respect to one-period ahead forecasts. Hence, MMN models are viable, if not superior, alternatives to modeling financial returns. Moreover, as evidenced in the post-predictive analyses of the GARCH and mixture models in Geweke and McCausland (2001), the mixture

Model	Series I (1981-1986)	Series II (1987-1991)	Series III (1981-1991)
GARCH	5,034.9	3,967.2	8,994.2
	(0.0716)[a]	(0.0838)	0.0935)
t-GARCH	5,062.8	4,062.5	9,140.2
	(0.1532)	(0.0962)	(0.0596)
MMN(2)	5,058.5	4,016.3	9,058.6
	(0.0518)	(0.0420)	(0.0402)
MMN(3)	5,060.9	4,051.7	9,113.1
	(0.4326)	(0.0777)	(0.0745)
MMN(4)	5,061.5	4,064.2	9,117.6
	(3.2742)	(3.2979)	(0.7711)

Note: [a]Numerical standard errors are in parantheses

Table VI. Log-marginal likelihood approximations models make much more precise statements about the characteristics of future observables, whereas the GARCH and t-GARCH models lead to quite vague predictions that are essentially inconsistent with what have been observed[8]. This strong contrast between the mixture models and the GARCH models have interesting and important asset pricing implications via these models. A full investigation of this problem is beyond the scope of this paper.

Given the encouraging performance of MMN models to modeling financial returns, it is hopeful that the MMN models will have success in other interesting financial applications such as VaR, and option pricing. Finally, in portfolio allocation and in risk management, a good model of the joint dynamics of multiple assets returns is very important to effective investment decision making. The univariate MMN structure can be generalized to a multivariate setting, which can provide a flexible yet practical approach to modeling multiple time series of assets returns that accounts for time-varying conditional distributions. Pursuit of this application remains for future research.

Notes

1. Formal tests have also been conducted to ensure that the MCMC outputs indeed are simulated correctly from the desired posterior distributions.
2. These overall statistics are what really matter, because in Bayesian framework what matters in prediction is $p(y_{T+1}|y_T, y_{T-1}, \ldots, y_1)$, the model parameters are integrated out.
3. I also tried diffuse (flat) priors, and the results change little from those reported here. This is not surprising because with large sample size, the likelihood dominates the priors in the posterior distribution.
4. Improper priors could drive the results of Bayesian model comparisons. For example, consider the hypothesis testing of model 1 versus model 2. If improper prior is used for model 1, then we could get the Bayes factor of 0/some constant. In this circumstance, the hypothesis test can never conclude in favor of model 1. This result is known as Lindley's paradox.
5. For the MMN(m) model, skewness can be easily addressed without additional structure. However, since GARCH and t-GARCH models have unconditional skewness equal zero by construction, we do not attempt to calibrate the priors for unconditional skewness. Additional structure on GARCH and t-GARCH models is needed if one desires to account for skewness in the prior calibrations.
6. For uniform prior beliefs over the set of models under consideration, the Bayes factor in favor of model i over model j is also the posterior odds ratio for the two models.
7. Since, in theory θ^* can be any value in Θ I take θ^* to be the θ associated with the highest posterior density throughout the analyses in this paper to increase computational efficiency.
8. For example, the quantile ratios (defined as the ratio of the range statistic to the interquartile range) implied by the MMN model and the GARCH (or t-GARCH) model are strikingly different.
9. See Chib (1996) for details on this.

References

Albert, J. and Chib, S. (1993), "Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts", *Journal of Business and Economic Statistics*, Vol. 11, pp. 1-15.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, Vol. 31, pp. 307-27.

Chib, S. (1995), "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, Vol. 90, pp. 1313-21.

Chib, S. (1996), "Calculating posterior distributions and modal estimates in Markov mixture models", *Journal of Econometrics*, Vol. 75, pp. 79-97.

Engle, R.F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation", *Econometrica*, Vol. 50, pp. 987-1008.

Engle, R.F. and Bollerslev, T. (1986), "Modelling the persistence of conditional variances", *Econometric Rev.*, Vol. 5, pp. 1-50.

Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration", *Econometrica*, Vol. 57 No. 6, pp. 1317-39.

Geweke, J. (1994), "Comment on Bayesian analysis of stochastic volatility", *Journal of Business and Economic Statistics*, Vol. 12 No. 4, pp. 371-417.

Geweke, J. and McCausland, W. (2001), "Bayesian specification analysis in econometrics", *American Journal of Agricultural Economics*, Vol. 83, pp. 1181-6.

Jacquier, E., Polson, N. and Rossi, P. (1994), "Bayesian analysis of stochastic volatility models", (with discussion), *Journal of Business and Economic Statistics*, Vol. 12 No. 4, pp. 371-417.

Lindgren, G. (1978), "Markov regime models for mixed distributions and switching regressions", *Scandinavian Journal of Statistics*, Vol. 5, pp. 81-91.

Ryden, T., Terasvirta, T. and Asbrink, S. (1998), "Stylized facts of daily return series and the hidden Markov model", *Journal of Applied Econometrics*, Vol. 13, pp. 217-44.

Further reading

Bollen, N., Gray, S. and Whaley, R. (2000), "Regime switching in foreign exchange rates: evidence from currency option prices", *Journal of Econometrics*, Vol. 94, pp. 239-76.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992), "ARCH modeling in finance", *Journal of Econometrics*, Vol. 52, pp. 5-59.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994) in Engle, R.F. and Mcfadden, D.L. (Eds), *Handbook of Econometrics*,Vol. IV, North-Holland, Amsterdam.

Bos, C., Mahieu, R. and Dijk, H. (2000), "Daily exchange rate behavior and hedging of currency risk", working paper, Erasmus University, Rotterdam.

Campbell, J., Lo, A.W. and MacKinley, A.C. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Carlin, J., Gelman, A., Rubin, D. and Stern, H. (1995), *Bayesian Data Analysis*, Chapman & Hall, New York, NY.

Chib, S. and Greenberg, E. (1995), "Understanding the metropolis-Hastings algorithm", *The American Statistician*, Vol. 49 No. 4, pp. 327-35.

Diebold, F.X. and Lopez, J.A. (1996), "Forecast evaluation and combination", *Handbook of Statistics*,Vol. 14, Elsevier Science, Amsterdam.

Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments", in Bernardo, J.M., Berger, J., David, A. and Smith, A. (Eds), *Bayesian Statistics*,Vol. 4, pp. 169-93.

Geweke, J. (1995), "Bayesian comparison of econometric models", Working Paper 532, Federal Reserve Bank of Minneapolis.

Geweke, J. (1999), "Using simulation methods for Bayesian econometric models: inference, development and communication", *Econometric Reviews*, Vol. 18, pp. 1-126.

Geweke, J. and Keane, M. (1999), "Mixture of normals probit models", in Hsiao, C., Lahiri, K., Lee, L-F. and Pesaran, M.H. (Eds), *Analysis of Panels and Limited Dependent Variables: In Honor of G.S. Maddala*, Cambridge University Press, Cambridge, pp. 49-78.

Hall, J., Brorsen, B. and Irwin, S. (1989), "The distribution of futures prices: a test of the stable paretian and mixture of normals hypotheses", *Journal of Financial and Quantitative Analysis*, Vol. 24 No. 1, pp. 105-16.

McDonald, J. (1996), "Probability distributions for financial models", *Handbook of Statistics*,Vol. 14, Elsevier Science, Amsterdam, pp. 427-61.

Vrontos, D., Dellaportas, P. and Politis, D.N. (2000), "Full Bayesian inference for GARCH and EGARCH models", *Journal of Business and Economic Statistics*, Vol. 18 No. 2, pp. 187-98.

Zellner, A. (1996), "Models, prior information, and Bayesian analysis", *Journal of Econometrics*, Vol. 75, pp. 51-68.

Appendix 1. Univariate Markov mixture of normals (MMN)

$$y_t|(\theta, s_t = i) \sim N\left(\mu_i, h_i^{-1}\right) \quad i = 1, \ldots, m \tag{A.1}$$

$$s_t|s_{t-1} \sim \text{Markov}(P, \pi_1)$$

Let $Y_T = \{y_1, \ldots, y_t\}$, $\phi_i = \{\mu_i, h_i\}$, and $\theta = \{\phi_1, \ldots, \phi_m, P, \pi_1\}$, then the normalized density is given by:

$$f(Y_T|\theta) = \prod_{t=1}^{T} f(y_t|Y_{T-1}, \theta) \tag{A.2}$$

where

$$f(y_t|Y_{T-1}, \theta) = \sum_{j=1}^{m} f(y_t|Y_{T-1}, \phi_j) p(s_t = j|Y_{t-1}, \theta) \tag{A.3}$$

Simulation of the states (Chib's Algorithm)[9]
Denote:

$$S_t = (s_1, \ldots, s_t), \quad S^{t+1} = (s_{t+1}, \ldots, s_n)$$

$$Y_t = (y_1, \ldots, y_t), \quad Y^{t+1} = (y_{t+1}, \ldots, y_n)$$

The joint density $s_1, s_1, \ldots, s_n|Y_n, \theta$ can be written as:

$$p(S_n|Y_n, \theta) = p(s_n|Y_n, \theta) \times \cdots \times p(s_t|Y_n, S^{t+1}, \theta) \times \cdots \times p(s_1|Y_n, S^2, \theta) \tag{A.4}$$

in which the typical term, excluding the terminal point, is given by:

$$p(s_t|Y_n, S^{t+1}, \theta)$$

By Bayes theorem:

$$
\begin{aligned}
p(s_t|Y_n, S^{t+1}, \theta) &\propto p(s_t|Y_t, \theta) \times f(Y^{t+1}, S^{t+1}|Y_t, s_t, \theta) \\
&\propto p(s_t|Y_t, \theta) \times p(s_{t+1}|s_t, \theta) \times f(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta) \\
&\propto p(s_t|Y_t, \theta) \times p(s_{t+1}|s_t, \theta)
\end{aligned} \tag{A.5}
$$

since the term $f(Y^{t+1}, S^{t+2}|Y_t, s_t, s_{t+1}, \theta)$ is independent of s_t.

To calculate $p(s_t|Y_t, \theta)$, assuming the function $p(s_{t-1}|Y_{t-1}, \theta)$ is available, then repeat the following steps:

- *Prediction step*. Determination of $p(s_t|Y_{t-1}, \theta)$. By the law of total probability:

$$
p(s_t|Y_{t-1}, \theta) = \sum_{k=1}^{m} p(s_t|s_{t-1} = k, \theta) \times p(s_{t-1} = k|Y_{t-1}, \theta) \tag{A.6}
$$

- *Update step*. Determination of $p(s_t|Y_t, \theta)$. By Bayes theorem, the mass function of the state given information up to time t is now:

$$
p(s_t|Y_t, \theta) \propto p(s_t|Y_{t-1}, \theta) \times f(y_t|Y_{t-1}, \theta_{s_t}) \tag{A.7}
$$

Prior specification

The priors used are based on the set of conditionally conjugate priors for the means, precisions and transition matrix:

$$
p(\mu_i) \sim N\left(\underline{\mu_i}, \underline{h_i}^{-1}\right) \tag{A.8}
$$

$$
p(h_i) \sim \text{Gamma}\left(\underline{\alpha_i}, \underline{\beta_i}\right) \tag{A.9}
$$

$$
p(P_i) \sim \text{Dirichlet}\left(\underline{c_{i1}}, \ldots, \underline{c_{im}}\right) \quad i = 1, \ldots, m \tag{A.10}
$$

Posterior distribution

Given the above priors and the data density, the posterior of θ can be expressed as:

$$
\begin{aligned}
p(\theta|Y_T) \propto \prod_{i=1}^{m} \exp\left[\frac{-\left(\mu_i - \underline{\mu_i}\right)^2}{2\underline{h_i}^{-1}}\right] \cdot h_i^{\underline{\alpha_i}-1} \exp\left(-\underline{\beta_i} h_i\right) \cdot \\
\prod_{i=1}^{m}\prod_{j=1}^{m} p_{ij}^{\underline{c_{ij}}-1} \cdot \\
\prod_{t=1}^{T} f(y_t|Y_{T-1}, \theta)
\end{aligned} \tag{A.11}
$$

Further, conditioning on the states $\{s_t\}$, the posterior above becomes:

$$
p(\theta|Y_T) \propto \prod_{i=1}^{m}\prod_{j=1}^{m} p_{ij}^{\underline{c_{ij}}-1} \cdot
$$

$$\prod_{i=1}^{m} \exp\left[\frac{-\left(\mu_i - \underline{\mu_i}\right)^2}{2\underline{h_i}^{-1}}\right] \cdot h_i^{\underline{\alpha_i}-1} \exp\left(-\underline{\beta_i} h_i\right) \cdot$$

$$\prod_{i=1}^{m} h_i^{n_i/2} \exp\left[\frac{-\sum_{k=1}^{n_i}\left(y_{i,k} - \mu_i\right)^2}{2h_i^{-1}}\right] \tag{A.12}$$

where $y_{i,k} = 1, \ldots, n_i$ are the observations assigned to state s_i

Based on the posterior, the Gibb's sampling is to run the following simulations successively:

$$s_t \sim s_1, s_2, \ldots, s_T | Y_T, \theta \quad t = 1, \ldots, T \tag{A.13}$$

$$\mu_i \sim N\left[\frac{\underline{h_i}\underline{\mu_i} + n_i h_i \overline{y_{n_i}}}{\underline{h_i} + n_i h_i}, \left(\underline{h_i} + n_i h_i\right)^{-1}\right] \quad i = 1, \ldots, m \tag{A.14}$$

$$h_i \sim \text{Gamma}\left[\underline{\alpha_i} + \frac{n_i}{2}, \underline{\beta_i} + \frac{\sum_{k=1}^{n_i}\left(y_{i,k} - \mu_i\right)^2}{2}\right] \quad i = 1, \ldots, m \tag{A.15}$$

$$p_i \sim \text{Dirichlet}\left(\underline{c_{i1}} + n_{i1}, \ldots, \underline{c_{im}} + n_{im}\right) \quad i = 1, \ldots, m \tag{A.16}$$

where

$$\overline{y_{n_i}} = \frac{\sum_{k=1}^{n_i} y_{i,k}}{n_i} \tag{A.17}$$

n_{ij} = total number of one-step transitions from state i to state j.

This process generates a Markov chain, which under mild conditions coverges under the L^1 norm to the desired posterior distribution. The output of the Markov chain, once it has passed its transient stage, is taken as a sample from the posterior distribution for purposes of computing moments and marginal densities.

Appendix 2. Unconditional statistics for MMN

Under the classical statistics setting, Ryden *et al.* (1998) provide the equations for the unconditional statistics of Markov mixture models as outlined below: In the general case of d states, the model for r_t can be written as:

$$r_t = \sum_{i=1}^{d} X_{it} \cdot I(S_t = i) \tag{B.1}$$

where $I(\cdot)$ is an indicator function, and $X_{it}, \quad i = 1, \ldots, d$, are independent normal variables.

Let:

$P = [p_{ij}]$ be the $(d \times d)$ matrix of transition probabilities

$$p_{ij}^{n} = \Pr(S_{t+n} = j|S_t = i) \text{ be the } n-\text{step transition probability}$$

$$P^{(n)} = P^{n} = \left[p_{ij}^{n}\right] \text{ be the corresponding } n-\text{step transition matrix}$$

$$\pi = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_d \end{bmatrix} \text{ be the stationary probabilities}$$

Let g be a function of the random variable r_t, and let:

$$G_i \equiv E[g(r_t)|S_t = i] \tag{B.2}$$

That is, G_i is the expected value of $g(r_t)$ given regime i. For instance, if $g(r_t) = r_t$, then G_i equals the conditional first moment. If $g(r_t) = r_t^2$, then G_i equals the conditional second moment. Generally, we have the following equations for the unconditional statistics of interest:

$$E[g(r_t)] = \sum_{i=1}^{d} \pi_i G_i = \pi G\mathbf{1} \quad \text{where} \quad G = \text{diag}(G_1 \ldots G_d), \quad \text{and} \quad \mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}' \tag{B.3}$$

$$\begin{aligned} E[g(r_t)g(r_{t+n})] &= \sum_{i=1}^{d}\sum_{j=1}^{d} E[g(r_t)g(r_{t+n})|S_t = i, S_{t+n} = j] \cdot \Pr(S_t = i, S_{t+n} = j) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} E[g(r_t)|S_t = i]E[g(r_{t+n})|S_{t+n} = j] \cdot \Pr(S_{t+n} = j|S_t = i) \cdot \Pr(S_t = i) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} G_i G_j p_{ij}^{n} \pi_i \quad = \pi G P^{(n)} G\mathbf{1} \\ Cov[g(r_t), g(r_{t+n})] &= \pi G P^{(n)} G\mathbf{1} - (\pi G\mathbf{1})^2 \end{aligned} \tag{B.4}$$

Hence, under the Bayesian framework, let $\mu^{(m)}, \sigma^{2(m)}, \text{Skew}^{(m)}, \text{Kurt}^{(m)}, \gamma^{(m)}$, and $\rho^{(m)}$ denote, respectively, the population mean, variance, skewness, kurtosis, first-order autocovariance, and first-order autocorrelation coefficient of the Markov mixture model computed based on the mth $(m = 1, \ldots, M)$ posterior draw and the equations defined in (IV.1) through (IV.5). Then, the following sample averages provide the estimates of the unconditional statistics under the Baysian framework:

$$\bar{\mu} = M^{-1} \sum_{m=1}^{M} \mu^{(m)} \tag{B.5}$$

$$\overline{\sigma^2} = M^{-1} \sum_{m=1}^{M} \sigma^{2(m)} \tag{B.6}$$

$$\overline{\text{Skew}} = M^{-1} \sum_{m=1}^{M} \text{Skew}^{(m)} \tag{B.7}$$

$$\overline{\text{Kurt}} = M^{-1} \sum_{m=1}^{M} \text{Kurt}^{(m)} \tag{B.8}$$

$$\bar{\gamma} = M^{-1} \sum_{m=1}^{M} \gamma^{(m)} \tag{B.9}$$

$$\bar{\rho} = M^{-1}\sum_{m=1}^{M}\rho^{(m)} \tag{B.10}$$

Appendix 3. GARCH models

The GARCH(1,1) model with Gaussian innovations (referred to as GARCH) for a univariate time series $\{y_t\}$ is,

$$y_t = \mu + \varepsilon_t$$

$$\varepsilon_t = \sqrt{h_t}\cdot v_t \tag{C.1}$$

$$h_t = \alpha + \delta\varepsilon_{t-1}^2 + \gamma h_{t-1} \tag{C.2}$$

$$v_t \overset{IID}{\sim} t(0,1)$$

$$\alpha > 0, \delta \geq 0, \gamma \geq 0, \delta + \gamma < 1.$$

Conditional on h_1, the likelihood function is:

$$L_T(\alpha,\gamma,\delta|Y_T) = \prod_{t=1}^{T}(2\pi h_t)^{-1/2}\exp\left[\frac{-(y_t-\mu)^2}{2h_t}\right] \tag{C.3}$$

The predictive density, through observation s, is:

$$(2\pi)^{-(s-t)/2}\prod_{t=T+1}^{s}h_t^{-1/2}\exp\left[\frac{-(y_t-\mu)^2}{2h_t}\right] \tag{C.4}$$

The following priors specifications as used in Geweke (1994) are employed for the GARCH model,

$$\mu \sim N\left(\underline{\mu}, \underline{\sigma}^2\right) \tag{C.5}$$

$$\log(\alpha) \equiv a \sim N\left(\underline{a}, \underline{s_a}^2\right) \tag{C.6}$$

$$\delta, \gamma, 1-\delta-\gamma \sim \mathrm{Dir}(\underline{c}) \tag{C.7}$$

Two posterior simulation algorithms are used to cross check the estimation results:

(1) Independent Metropolis-Hastings method

The computations for the posterior can be performed by constructing a Metropolis independence chain to produce a sequence of parameters whose unconditional limiting distribution is the posterior distribution, as suggested in Geweke (1994).

Let $\theta' = (\mu, a, \delta, \gamma)$, and let $p_T(\theta|Y_T)$ denote the posterior distribution at time T. Denote the mode of the (log) posterior kernel by $\hat{\theta}$, and the Hessian at the mode by $\mathbf{H}$. Let $J(\cdot; \mu, \mathbf{V}, \nu)$denote the kernel density of a multivariate Student-t distribution with location vector μ, scale matrix $\mathbf{V}$, and ν degrees of freedom. For the choices $\mu = \hat{\theta}$, $\mathbf{V} = -(1.2)^2\mathbf{H}^{-1}$, $\nu = 5$, the ratio $(p_T(\theta|Y_T))/(J(\theta; \mu, \mathbf{V}, \nu))$ is bounded above.

This multivariate Student-t distribution forms a proposal distribution for an independece Metropolis algorithm as follows. At step m, generate a candidate θ^* from $J(\cdot;\mu,\mathbf{V},\nu)$. Then:

$$\begin{cases} \theta^{(m)} = \theta^* & \text{, with probability } p. \\ \theta^{(m)} = \theta^{(m-1)} & \text{, with probability } 1-p. \end{cases}$$

$$\text{where} \quad p = \min\left\{\frac{p_T(\theta^*|Y_T)/J(\theta^*;\mu,\mathbf{V},\nu)}{p_T(\theta^{(m-1)}|Y_T)/J(\theta^{(m-1)};\mu,\mathbf{V},\nu)}, 1\right\} \tag{C.8}$$

(2) Random-walk Metropolis method

The second posterior simulation algorithm employed for this model is the random-walk variant of the Hastings-Metropolis algorithm, (random-walk Metropolis). The general Hastings-Metropolis algorithm generates candidate draws θ^* from a source density kernel q given the current value $\theta^{(m)}$, and sets $\theta^{(m+1)} = \theta^*$ with probability,

$$\alpha(\theta^{(m)},\theta^*) = \begin{cases} \min\left\{\frac{p(\theta^*|Y_T)q(\theta^*,\theta^{(m)})}{p(\theta^{(m)}|Y_T)q(\theta^{(m)},\theta^*)}, 1\right\} & \text{if } p(\theta^{(m)}|Y_T)q(\theta^{(m)},\theta^*) > 0 \\ 1 & \text{if } p(\theta^{(m)}|Y_T)q(\theta^{(m)},\theta^*) = 0 \end{cases} \tag{C.9}$$

The random-walk variant generates ε independently from the density f, and then sets $\theta^* = \theta^{(m)} + \varepsilon$. Thus, $q(\theta^{(m)}, \theta^*) = f(\theta^* - \theta^{(m)})$.

For this particular case, f was taken to be the multivariate Student-t density with mean 0, η degrees of freedom, and precision matrix λ. In each case, λ was set to $-c^{-2}H$, where H is the approximate Hessian of the log posterior at the posterior mode, and c is a step parameter. Since, the multivariate Student-t is symmetric, $q(\theta^*, \theta^{(m)}) = q(\theta^{(m)}, \theta^*)$, so the acceptance probability becomes:

$$\alpha(\theta^{(m)},\theta^*) = \begin{cases} \min\left\{\frac{p(\theta^*|Y_T)}{p(\theta^{(m)}|Y_T)}, 1\right\} & \text{if } p(\theta^{(m)}|Y_T)q(\theta^{(m)},\theta^*) > 0 \\ 1 & \text{if } p(\theta^{(m)}|Y_T)q(\theta^{(m)},\theta^*) = 0 \end{cases} \tag{C.10}$$

Approximating the (log) marginal likelihood of the GARCH models

There are at least two possible approaches to obtaining an approximation of the (log) marginal likelihood that do not require additional simulation or a distinct simulation algorithm. I use the most basic approach which is based upon the fact that for any density g whose support contains Θ,

$$M(Y_T) = E_g\left[\frac{p(\theta)p(Y_T|\theta)}{g(\theta)}\right] \tag{C.11}$$

Therefore, if the candidate draws θ^* have the density g, an approximation is obtained by evaluating $p(\theta^*)p(Y_T|\theta^*)/g(\theta^*)$ for each θ^* then compute the sample analog of the above equation (Geweke, 1989).

The GARCH(1,1) model with Student-t innovations (referred to as t-GARCH) for a univariate time series $\{y_t\}$ is,

$$y_t = \mu + \varepsilon_t \tag{C.12}$$

$$\varepsilon_t = \sqrt{h_t}\cdot v_t \tag{C.13}$$

$$h_t = \alpha + \delta\varepsilon_{t-1}^2 + \gamma h_{t-1} \tag{C.14}$$

$$v_t \overset{IID}{\sim} t(0, s, \nu)$$

$$\alpha > 0, \delta \geq 0, \gamma \geq 0, \delta + \gamma < 1.$$

where the scale parameter s is set equal to $(v_-2)/v$, so that $\mathrm{var}(v_t) = 1$. Note that $y_t t(\mu,((v-2)/v)h_t,\nu)$, and $\mathrm{var}(y_t) = h_t$. The condition that $\nu > 2$ is imposed in the prior.

Conditional on h_1, the likelihood function is:

$$L_T(\alpha, \gamma, \delta, \nu|Y_T) = \prod_{t=1}^{T} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)}[(v-2)\pi h_t]^{-1/2}\left[1 + \frac{(y_t - \mu)^2}{(v-2)h_t}\right]^{-(v+1)/2} \tag{C.15}$$

The predictive density, through observation s, is:

$$\pi^{-(s-T)/2} \prod_{t=T+1}^{s} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)}[(v-2)h_t]^{-1/2}\left[1 + \frac{(y_t - \mu)^2}{(v-2)h_t}\right]^{-(v+1)/2} \tag{C.16}$$

The following priors are employed for the t-GARCH model,

$$\mu \sim N(\underline{\mu}, \underline{\sigma^2}) \tag{C.17}$$

$$\log(\alpha) \equiv a \sim N\left(\underline{a}, \underline{s_a}^2\right) \tag{C.18}$$

$$\delta, \gamma, 1 - \delta - \gamma \sim \mathrm{Dir}(\underline{c}) \tag{C.19}$$

$$v - 4 \sim \chi^2(\underline{v}) \tag{C.20}$$

The usual nonnegativity restrictions are imposed in the prior. The same two posterior simulation algorithms used for the GARCH(1,1) model are employed for the t-GARCH model, with appropriate modifications.

Corresponding author
George Chang can be contacted at: ygchang@bradley.edu