

Bayesian analysis of financial volatilities addressing long-memory, conditional heteroscedasticity and skewed error distribution

Rosy Oh^a, Dong Wan Shin^a, Man-Suk Oh^{1,a}

^aDepartment of Statistics, Ewha Womans University, Korea

Abstract

Volatility plays a crucial role in theory and applications of asset pricing, optimal portfolio allocation, and risk management. This paper proposes a combined model of autoregressive moving average (ARFIMA), generalized autoregressive conditional heteroscedasticity (GRACH), and skewed- t error distribution to accommodate important features of volatility data; long memory, heteroscedasticity, and asymmetric error distribution. A fully Bayesian approach is proposed to estimate the parameters of the model simultaneously, which yields parameter estimates satisfying necessary constraints in the model. The approach can be easily implemented using a free and user-friendly software JAGS to generate Markov chain Monte Carlo samples from the joint posterior distribution of the parameters. The method is illustrated by using a daily volatility index from Chicago Board Options Exchange (CBOE). JAGS codes for model specification is provided in the Appendix.

Keywords: ARFIMA, GARCH, skewed- t , JAGS, Bayesian, Markov chain Monte Carlo

1. Introduction

Volatility plays a crucial role in the theory and application of asset pricing, optimal portfolio allocation, and risk management. It is therefore important to introduce a model addressing the key features of volatility data sets and develop a valid estimation method. The most dominant features of the usual volatilities, the historic volatilities, the implied volatilities, and the realized volatilities are long memories, conditional heteroscedasticity and skewed error distribution.

The fractionally integrated autoregressive moving average (ARFIMA) model (Granger and Joyeux, 1980; Hosking, 1981) and the generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986) have been widely used to explain the long memory and the time-varying conditional variance features of volatilities, respectively. Park (2016) used a skewed student t distribution to incorporate asymmetry as well as fat tails of error distribution (Corsi *et al.*, 2008; Grassi and Magistris, 2015; Graves *et al.*, 2015; Maasoumi and McAleer, 2008; Wang *et al.*, 2013).

However, none of the currently available methods propose a model incorporating all the above three features simultaneously. For example, Baillie *et al.* (1996) proposed the ARFIMA + GARCH model but did not incorporate the asymmetry of the error distribution. Park (2016) proposed a heterogeneous autoregressive (HAR) model to incorporate the long memory and then employed a GARCH + skewed- t distribution model to the estimated residuals obtained from fitting the HAR model to

¹ Corresponding author: Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: msoh@ewha.ac.kr

volatility data sets. Consequently, they incorporated the three features of volatilities in a stepwise fashion rather than simultaneously.

In this paper, we propose a model which combines ARFIMA, GARCH, and skewed- t distribution, namely ARFIMA + GARCH + skewed- t model, to simultaneously account for the long memory, time-varying variance, and asymmetric error distribution features of volatilities.

A Bayesian approach is used to estimate the parameters of the model. We employ Just Another Gibbs Sampler (JAGS, <http://mcmc-jags.sourceforge.net/>) to generate Markov chain Monte Carlo (MCMC) samples from the joint posterior distributions of the parameters. A key advantage of JAGS is that the software runs MCMC automatically once a user specifies a model (the distribution of data and the prior distributions of parameters); therefore, practitioners who are not experts in MCMC may also easily implement Bayesian inference using JAGS.

Given the posterior samples, several features of interest, such as the estimated marginal posterior densities, posterior moments, quantiles, scatter plots exhibiting interesting relations between parameters, may be derived in a straightforward manner. In addition, any restrictions on parameters, such as inequality restrictions on coefficients in the GARCH model to ensure stationarity, can be implemented in a straightforward way. Finally, useful prior information, which may be available from the data context or previous data analysis, can be used in Bayesian approaches (Czado and Min, 2011).

This paper is organized as follows. Section 2 presents the ARFIMA + GARCH + skewed- t model and the Bayesian estimation method. Section 3 presents analysis of a real data set. The final section draws conclusions.

2. Method

In this section, we describe the proposed model, the priors for the parameters, and a Markov chain Monte Carlo method for Bayesian inference.

2.1. Model

We consider a model that captures the features of conditional heteroscedasticity, long-memory, asymmetry, and fat tails via a combination of ARFIMA, GARCH, and skewed- t distribution. Long memories of volatility process y_t is modeled by the ARFIMA(p, d, q) process

$$\Phi(B)(1 - B)^d y_t = \theta_0 + \Theta(B)e_t, \quad (2.1)$$

where e_t is a zero mean uncorrelated error sequence, $d \in (-0.5, 1)$ is the fractional integration parameter, B is the backward-shift operator, and $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ are assumed to have roots outside the unit circle and have no common roots. The fractional differencing operator, $(1 - B)^d$, is defined as $(1 - B)^d = \sum_{k=0}^{\infty} \pi_k B^k$, where $\pi_0 = 1$, and $\pi_k = \pi_{k-1}(k - 1 - d)/k$ for $k > 0$. If $d \in (-0.5, 0.5)$, y_t is stationary; if $d \in [0.5, 1)$, y_t is nonstationary. For nonstationary series, Bayesian analysis can be made on the differenced series $w_t = (1 - B)y_t$, which is always stationary for $d \in (-0.5, 1]$. The difference series w_t is an ARFIMA(p, δ, q) given by $\Phi(B)(1 - B)^\delta (w_t - \mu) = \Theta(B)e_t$, where $\delta = d - 1$ and $\mu = E(w_t)$.

From Bollerslev *et al.*, (1992), in general a GARCH(1, 1) model is sufficient to capture volatility clustering in the data. Therefore, in order to address the time-varying variance of y_t , the error e_t in (2.1) can be assumed to follow a GARCH(1, 1) model given by

$$e_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where ε_t is a sequence of i.i.d. errors having zero mean and unit variance. Positivity of σ_t^2 and stationarity of e_t are ensured via the standard restrictions: $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$.

Asymmetry and fat tails are featured for the zero-mean-unit-variance error ε_t to have a standardized skewed- t distribution with skewness parameter θ and degrees of freedom ν (Lambert and Laurent, 2001). It has a probability density function (pdf)

$$f(\varepsilon_t|\theta, \nu) = \begin{cases} \frac{2s}{\theta + \theta^{-1}} t_\nu(\theta(s\varepsilon_t + m)), & \text{if } \varepsilon_t < -\frac{m}{s}, \\ \frac{2s}{\theta + \theta^{-1}} t_\nu\left(\frac{s\varepsilon_t + m}{\theta}\right), & \text{if } \varepsilon_t \geq -\frac{m}{s}, \end{cases} \tag{2.2}$$

where m and s^2 are the mean and the variance of the nonstandardized skewed- t distribution; in addition, $t_\nu(\cdot)$ denotes the standard Student t density with ν degrees of freedom. The expected value and the variance of ε_t are given as $E(\varepsilon_t) = 0, V(\varepsilon_t) = 1$. The parameter θ indicates the direction and the degree of skewness. If $\theta = 1$, the skewed- t distribution reduces to the Student t distribution.

Now we consider a model consisting of ARFIMA(1, d , 0), GARCH(1, 1), and skewed- t distribution, for which we have

$$\begin{aligned} (1 - \phi B)(1 - B)^\delta(w_t - \mu) &= e_t, \\ e_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \\ \varepsilon_t &\stackrel{\text{iid}}{\sim} \text{Std.skt}(\theta, \nu), \end{aligned} \tag{2.3}$$

with constraints $\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1$. We will refer this model as ARFIMA(1, d , 0) + GARCH(1, 1) + skewed- t model. The combined model takes into account all the previously mentioned important features of volatility data sets.

2.2. Priors

We assume prior independence among $\mu, \phi, \delta, (\alpha_0, \alpha_1, \beta_1), \theta, \nu$. All parameters except for μ should satisfy some constraints. We incorporate these constraints in prior distributions of the parameters so that posterior samples generated from MCMC as well as estimates of the parameters automatically satisfy the constraints.

A diffuse normal prior is used for μ , i.e., $\mu \sim N(\mu_\mu, \sigma_\mu^2)$, with σ_μ large. We assume $\phi \sim N(\mu_\phi, \sigma_\phi^2)I(-1, 1)$ and $\delta \sim N(\mu_\delta, \sigma_\delta^2)I(-1, 0)$. For the prior of $(\alpha_0, \alpha_1, \beta_1)$, we assume $\pi(\alpha_0, \alpha_1, \beta_1) = \pi(\alpha_0, \alpha_1)\pi(\beta_1)I(\alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0, \alpha_1 + \beta_1 < 1)$, where $I(\cdot)$ is the indicator function. We choose $\pi(\beta_1)$ as $N(\mu_{\beta_1}, \sigma_{\beta_1}^2)$. For $\pi(\alpha_0, \alpha_1)$, we choose a hierarchical prior suggested by Ardia and Hoogerheide (2009):

$$(\alpha_0, \alpha_1) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}\right),$$

$\rho \sim N(\mu_\rho, \sigma_\rho^2)I(-1 < \rho < 1)$, $\mu_\rho \sim \text{Unif}(-1, 1)$, $\sigma_\rho^2 \sim \text{Gamma}(1.5, 10^{-4})$. Finally, we assume $\theta \sim N(\mu_\theta, \sigma_\theta^2)I(\theta > 0)$, and $\nu \sim N(\mu_\nu, \sigma_\nu^2)I(\nu \geq 2)$.

2.3. Estimation

Clearly, the posterior distribution of the parameters $\Theta = (\mu, \delta, \phi, \alpha_0, \alpha_1, \beta_1, \theta, \nu)$ in the model (2.3) is not analytically tractable. Simple MCMC methods such as standard Gibbs sampling methods are

not applicable to this model; therefore, Metropolis-Hastings (MH) algorithm is required to generate posterior samples of the parameters.

Implementation of the MH algorithm is often complicated for practitioners who are not experts in MCMC and/or coding. We get around this difficulty by generating MCMC samples by using user-friendly JAGS software. Given the model specification, it employs either Gibbs sampler when available or MH algorithm to generate posterior samples.

To use JAGS, the distribution of data needs to be specified as one of the distributions built in JAGS. However, the distributions in the ARFIMA(1, d , 0) + GARCH(1, 1) + skewed- t model are not commonly used distributions and are not provided by JAGS. Instead of specifying the data distribution, we can specify the likelihood function since JAGS derives the likelihood from the distribution of data and uses them for the Gibbs sampler or MH algorithm.

For the likelihood of Θ , we first derive the joint pdf of $e = (e_1, e_2, \dots, e_T)$ from the GARCH(1, 1) and the skewed- t models, which is given as

$$\begin{aligned} f(e_1, e_2, \dots, e_T | \Theta) &= f(e_1 | \Theta) f(e_2 | e_1, \Theta) \cdots f(e_T | e_1, e_2, \dots, e_{T-1}, \Theta) \\ &= f_\varepsilon\left(\frac{e_1}{\sigma_1}\right) \frac{1}{\sigma_1} f_\varepsilon\left(\frac{e_2}{\sigma_2}\right) \frac{1}{\sigma_2} \cdots f_\varepsilon\left(\frac{e_T}{\sigma_T}\right) \frac{1}{\sigma_T}, \end{aligned} \quad (2.4)$$

where $f_\varepsilon(\cdot)$ is the density function of standardized skewed- t distribution given in (2.2). Now we plug in the residuals in the ARFIMA model into (2.4). If w_t is invertible, the residual e_t in the ARFIMA model can be represented as an infinite autoregressive (AR) process as;

$$e_t = \sum_{k=0}^{\infty} \pi_k (w_{t-k} - \mu) - \phi (w_{t-k-1} - \mu) = \sum_{k=0}^{\infty} \psi_k (w_{t-k} - \mu), \quad (2.5)$$

where $\psi_k = \pi_k - \pi_{k-1}\phi$, $\psi_0 = 1$, $\pi_k = \pi_{k-1}(k-1-d)/k$ and $\pi_0 = 1$. The equation (2.5) is an infinite AR representation of ARFIMA(1, $d-1$, 0). However, when T is sufficiently large and the first difference w_t is stationary, we may replace (2.5) by a finite AR representation

$$e_t = \sum_{k=0}^{t-1} \psi_k (w_{t-k} - \mu). \quad (2.6)$$

In usual financial volatility analysis, since data sets are recorded daily or more frequently and the series length is large, this approximation would be good. Now the likelihood function of Θ in the ARFIMA(1, d , 0) + GARCH(1, 1) + skewed- t model is given by (2.4) with e_t , $t = 1, \dots, T$, given in (2.6).

The likelihood function (2.4) can be specified in the model specification of JAGS, using the method suggested by Ntzoufras (2011) and Kruschke (2014). The Appendix contains the model specification used in JAGS.

3. Analysis of the volatility index data

3.1. Exploratory study

We consider a daily closing price for the volatility index (VIX), which is a measure of market expectations on volatility over the next 30 days conveyed by S&P 500 stock index option prices. The data set can be obtained from the Chicago Board Options Exchange (CBOE) web site, <http://www.cboe.com/>

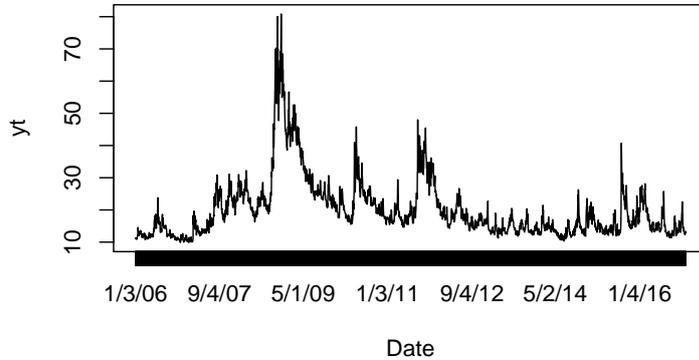


Figure 1: Time series plots of the volatility index.

Table 1: Summary statistics of the volatility index data

Min	Median	Mean	Max	SD	Skewness	Kurtosis
9.890	17.230	20.060	80.860	9.574	2.392	7.546

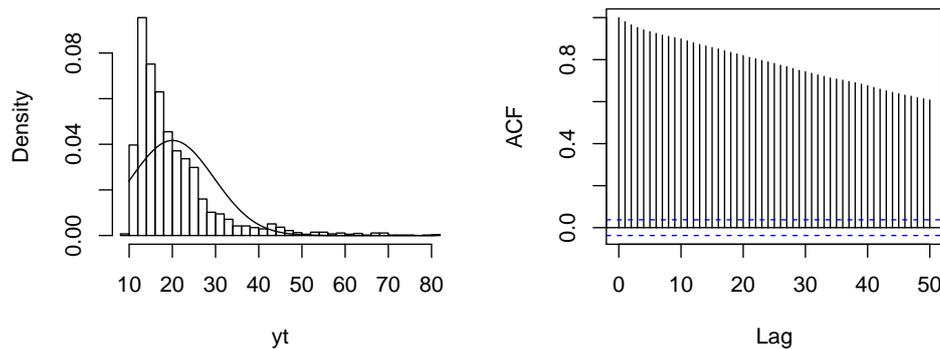


Figure 2: Histogram with normal density curve and ACF of the VIX. ACF = autocorrelation function; VIX = volatility index.

[micro/vix/historical.aspx](http://micro.vix/historical.aspx). We use data from January 3, 2006 to November 30, 2016, which contains 2,748 observations.

Time series plot of the VIX is presented in Figure 1. From the figure, we can see apparent conditional heteroscedastic feature of the VIX data; that high peaks from August 2007 to March 2009 due to the financial crisis of 2007–2008 cause large volatilities and that the relatively lower index values between 2013 and 2015 have smaller volatilities.

Table 1 shows basic summary statistics of the VIX. Figure 2 displays the histogram (with a normal approximation) and the autocorrelation function (ACF) plot of the VIX. The values of skewness, median, and mean (Table 1) and the histogram (Figure 2) show that the VIX is skewed to the right and has highly fat tails. The ACF plot of the VIX shows the feature of long memory which can be seen that the VIX is skewed to the right and has highly fat tails. The ACF plot of the VIX shows the feature of long memory which can be also supported by the GPH (Geweke and Porter-Hudak, 1983) test statistics of 8.663.

Table 2: Test for unit root

	ADF test	KPSS test
p -value	0.0422	< 0.01

ADF = augmented Dicky-Fuller; KPSS = Kwiatkowski-Phillips-Schmidt-Shin.

Table 3: Estimation results for the first difference of volatility index from ARFIMA + GARCH + skewed- t model

	Estimated posterior mean	Standard error	95 % credible interval	
			Lower	Upper
μ	-0.008	0.002	-0.013	-0.004
ϕ	0.186	0.036	0.118	0.256
δ	-0.332	0.029	-0.387	-0.276
α_0	0.068	0.014	0.046	0.099
α_1	0.213	0.023	0.170	0.260
β_1	0.778	0.022	0.732	0.818
θ	1.376	0.037	1.304	1.450
ν	4.419	0.317	3.871	5.113

ARFIMA = autoregressive moving average; GARCH = generalized autoregressive conditional heteroscedasticity.

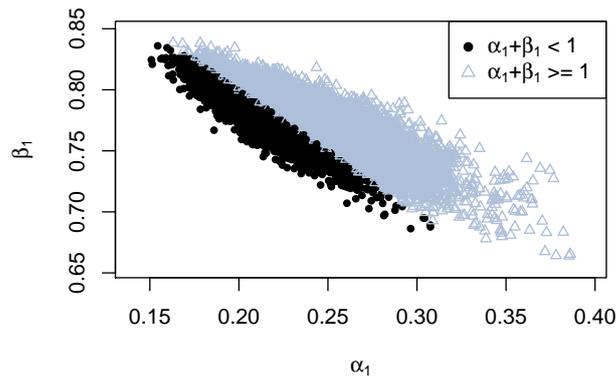
Figure 3: Scatter plot for α_1 and β_1 .

Table 2 presents the p -values of augmented Dicky-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test on the VIX (Dickey and Fuller, 1979; Kwiatkowski *et al.*, 1992). Null hypotheses of both the ADF and the KPSS tests are rejected at 5% level. Recall that the ADF and the KPSS tests have reversed the null hypothesis of unit root stationarity, respectively. It presumably indicates that there exists significant evidence that the process is a neither nonstationary nor stationary process but near nonstationary. To meet stationarity property of the series, we use the first differences of the VIX for analysis. Note that the differenced series is always stationary for $d \in (-0.5, 1]$.

3.2. Analysis

We applied the proposed Bayesian method to the first differences of the VIX data and generated posterior samples of parameters using JAGS. The first 3,000 iterations are discarded as burn-in and we obtained 15,000 samples after the burn-in from 3 different chains. Convergence of MCMC is diagnosed by Gelman-Rubin shrink factors and trace plots of the MCMC samples.

Table 3 displays the estimated posterior mean, standard error, and an approximate 95% credible

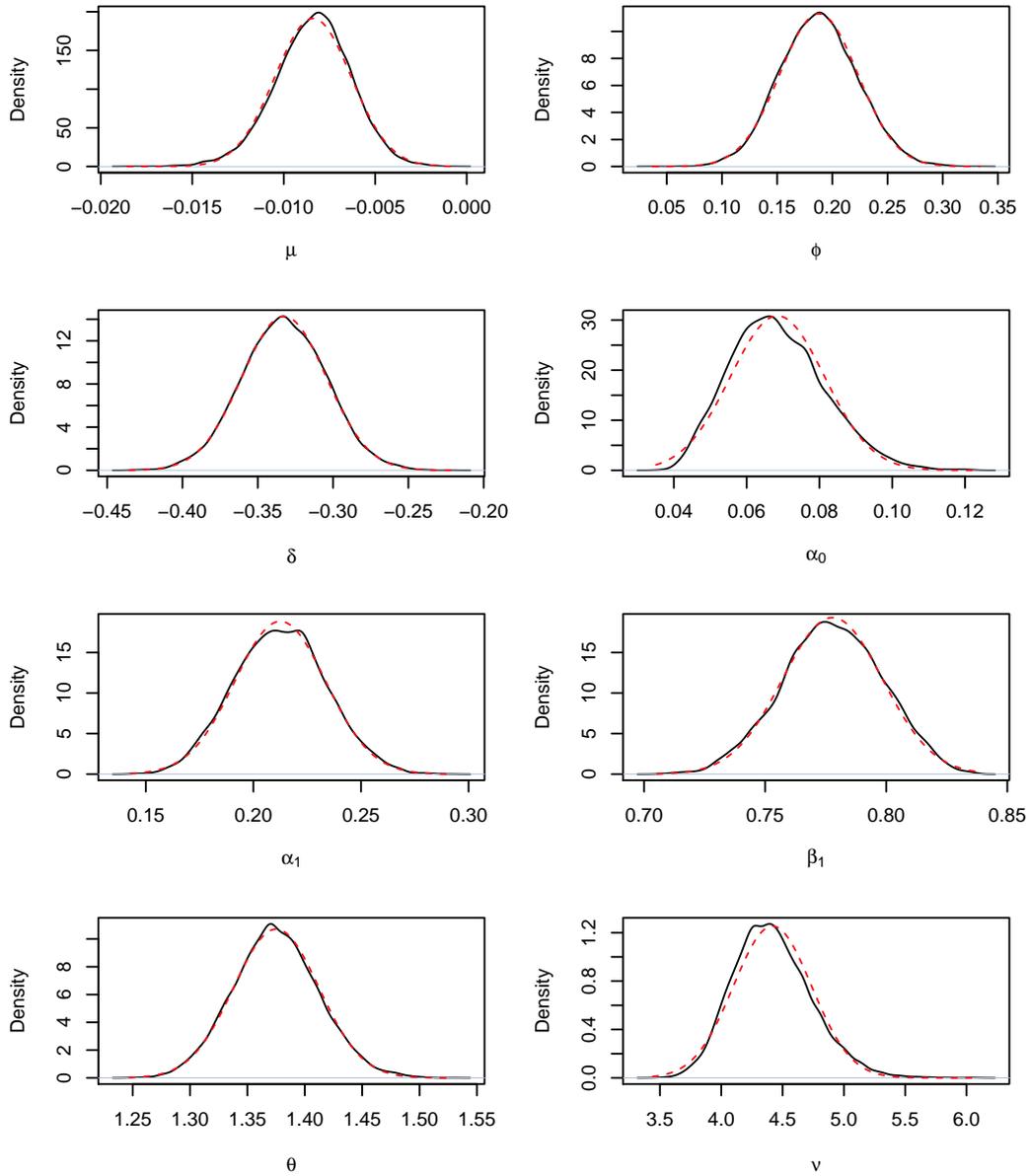


Figure 4: Estimated posterior distribution (solid line) and normal distribution (dashed line).

interval of each parameter. All parameters are significant in that their 95% credible intervals do not contain zero. The estimated posterior mean -0.332 of the fractional integration parameter δ indicates that process of the VIX has a nonstationary memory of fractional integration order of 0.668 . The sum $\hat{\alpha}_1 + \hat{\beta}_1 = 0.991$ indicates strong conditional heteroscedasticity. We also see significant asymmetry of the error distribution from the highly significant estimates $\hat{\theta} = 1.376$ of the skewness parameter.

The estimates satisfy the stationary condition $\hat{\alpha}_1 + \hat{\beta}_1 < 1$. Figure 3 displays the scatter plot of the

Table 4: Estimation results for the first difference of volatility index from various models

	95 % credible interval											
	ARFIMA + GARCH + skewed- t				ARFIMA + GARCH + t				ARFIMA + GARCH + normal			
	Est.	SE	Lower	Upper	Est.	SE	Lower	Upper	Est.	SE	Lower	Upper
μ	0.001	0.010	-0.018	0.020	-0.021	0.004	-0.029	-0.015	-0.013	0.004	-0.021	-0.006
ϕ	0.028	0.030	-0.031	0.088	0.189	0.044	0.105	0.278	0.155	0.046	0.070	0.253
δ	-0.181	0.024	-0.227	-0.133	-0.305	0.037	-0.379	-0.236	-0.263	0.038	-0.343	-0.192
σ^2	3.504	0.096	3.320	3.697	-	-	-	-	-	-	-	-
α_0	0.095	0.017	0.065	0.132	0.097	0.019	0.062	0.136	0.143	0.018	0.110	0.180
α_1	0.267	0.025	0.223	0.319	0.249	0.026	0.193	0.297	0.253	0.024	0.208	0.304
β_1	0.727	0.024	0.675	0.769	0.743	0.026	0.696	0.798	0.718	0.022	0.675	0.760
θ	1.287	0.030	1.230	1.346	-	-	-	-	-	-	-	-
ν	4.471	0.315	3.915	5.156	3.921	0.238	3.505	4.435	-	-	-	-

ARFIMA = autoregressive moving average; GARCH = generalized autoregressive conditional heteroscedasticity; Est. = Estimated posterior mean; SE = standard error.

Table 5: Deviance information criterion

skew- t	t	Normal
8886.260	9033.838	9508.023

samples of (α_1, β_1) when the constraints are ignored, in which the dark points represent the sample points satisfying the condition and white points, 69.54% of samples, represent those not satisfying the condition. The plot indicates that enforcing the constraints in the priors plays an important role to obtain the estimates within the stationarity region.

Figure 4 shows the estimated marginal posterior distributions of the parameters (solid line) with normal approximations (dashed line). The figure shows that marginally the posterior distributions of the parameters seem to be well approximated by normal distributions, except for α_0 and ν whose distributions show slight deviations from the normal approximations. Note the marginal posterior distributions are close to normal; however, the joint posterior distributions may be significantly different from multivariate normal (Figure 3).

For comparison, we applied a model ARFIMA + GARCH + skewed- t , which fits ARFIMA model to the data and then fit GARCH + skewed- t model to the estimated residuals from the ARFIMA model. The results are presented in Table 4.

The most notable differences are the estimates of the parameters of the ARFIMA model. In the ARFIMA + GARCH + skewed- t model, δ is much smaller and ϕ is much larger than those in the step-wise ARFIMA + GARCH + skewed- t model. This reveals that the heteroscedasticity and/or skewed error has affected the parameters of the ARFIMA model and hence the importance of simultaneous consideration of the three important features of volatilities.

We also applied ARFIMA + GARCH + t and ARFIMA + GARCH + normal models to the data, assuming t and normal error distributions, respectively, instead of skewed- t . Table 5 provides the Deviance Information Criteria (DIC) (Spiegelhalter *et al.*, 2002) for ARFIMA + GARCH + skewed- t , ARFIMA + GARCH + t , and ARFIMA + GARCH + normal models. Among the three models, ARFIMA + GARCH + skewed- t model is the best, having the smallest DIC, which reveals the asymmetry and heavy-tails of the error distribution. Table 4 provides estimation results from ARFIMA + GARCH + t and ARFIMA + GARCH + normal models. Compared with a skewed- t distribution for the error, assuming a t distribution yields a larger δ and a smaller degrees of freedom, and a normal distribution yields a much smaller ϕ and larger δ , α_0 and α_1 .

4. Conclusion

In this paper, we have proposed a fully Bayesian implementation of the ARFIMA + GARCH + skewed- t model which simultaneously takes into account the features of long memory, conditional heteroscedasticity, asymmetry and fat tails of volatility data sets. JAGS are used to generate MCMC samples of the parameters from their joint posterior distributions since it can be run automatically given model specification without expert knowledge on MCMC and coding.

Bayesian method naturally produce estimates that satisfy the stationary constraints in GARCH (1, 1), by incorporating constraints in the prior. The scatter plot of the samples of (α_1, β_1) show that restricting the parameter space is necessary to obtain parameter estimates that satisfy the constraints.

In this Bayesian implementation, since the likelihood function of the parameters of the ARFIMA + GARCH + skewed- t model is specified, modifying the model specification and choosing flat priors yield approximate conditional sum of squares (CSS) estimates within the stationary region based on MCMC samples. One may also apply an optimization method directly to the given likelihood function to obtain CSS estimates, instead of using MCMC samples. However, it is often relatively difficult to incorporate parameter constraints in optimization than in MCMC.

Acknowledgements

This research was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2016R1A2B4008914) (Rosy Oh and Man-Suk Oh) and (2016R1A2B4008780) (Dong Wan Shin).

Appendix A: Model specification in JAGS

```

data{
  zero <- 0
}
model {
  e[1] = w[1]-mu
  sig[1] = sig0
  epsilon[1]=e[1]/sig[1]

  Z[1] ~ dpois(-lnfi[1])
  lnfi[1] = ifelse(s*epsilon[1]+m<0, lnf1[1], lnf2[1])
  lnf1[1] = (log(2*s)-log(theta+ 1/theta)+
            loggam((nu+1)/2)-(1/2)*log(PI*(nu-2))-loggam(nu/2)-
            ((nu+1)/2)*log(1+((theta*(s*epsilon[1]+m))^2)/(nu-2)))-
            log(sig[1])
  lnf2[1] = (log(2*s)-log(theta+ 1/theta)+
            loggam((nu+1)/2)-(1/2)*log(PI*(nu-2))-loggam(nu/2)-
            ((nu+1)/2)*log(1+((s*epsilon[1]+m)/theta)^2)/(nu-2)))-
            log(sig[1])

  psi0 = 1
  psi1[1]= (-1)*phi + (-1)*d
  for(t in 2:n){
    Z[t] ~ dpois(-lnfi[t])
  }
}

```

```

lnfi[t] = ifelse(s*epsilon[t]+m<0, lnfi1[t], lnfi2[t])
lnfi1[t] = log(2*s)-log(theta+ 1/theta) +
          loggam((nu+1)/2)-(1/2)*log(PI*(nu-2))-loggam(nu/2)-
          ((nu+1)/2)*log(1+((theta*(s*epsilon[t]+m))^2)/(nu-2))-
          log(sig[t])

lnfi2[t] = log(2*s)-log(theta+ 1/theta)+
          loggam((nu+1)/2)-(1/2)*log(PI*(nu-2))-loggam(nu/2)-
          ((nu+1)/2)*log(1+((s*epsilon[t]+m)/theta)^2)/(nu-2))-
          log(sig[t])

psi1[t] = pi1[t]-pi1[t-1]*phi
for(k in 1:(t-1)){
  rev.psi1[t,k] = psi1[t-k]
}
epsilon[t] = e[t]/sig[t]
e[t] = psi0*(w[t]-mu) + inprod(rev.psi1[t,1:(t-1)], (w[1:(t-1)]-mu))
sig[t] = sqrt(alpha0 + alpha1*(e[t-1])^2 + beta1*(sig[t-1]^2))
}
pi0 = 1
pi1[1] = (-1)*d
for(k in 2:n){
  pi1[k] = pi1[k-1]*(k-1-d)/k
}
s = sqrt( theta^2 + theta^(-2) -1 - m^2 )
m = exp( loggam((nu-1)/2) + (1/2)*log(nu-2) - (1/2)*log(PI) -
        loggam(nu/2))*(theta - 1/theta)

## prior for ARFIMA(1,d,0)
mu ~ dnorm(m.mu, (s2.mu*100)^(-1))
phi ~ dnorm(m.phi, (s2.phi*100)^(-1))T(-1,1)
d ~ dnorm(m.d, (s2.d*100)^(-1))T(-1, 0)
PI = 3.14159265359

# prior for GARCH(1,1)
zero ~ dinterval(alpha1+beta1, 1)

alpha0 ~ dnorm(m.alpha0, 1/s2.A[1,1])T(0, )
alpha1 ~ dnorm(m.con.alpha1, inv.s2.con.alpha1)T(0, 1)
m.con.alpha1 = m.alpha1+s2.A[2,1]*(1/s2.A[1,1])*(alpha0-m.alpha0)
inv.s2.con.alpha1 = (s2.A[2,2]-(s2.A[1,2])*(1/s2.A[1,1])*(s2.A[2,1]))^(-1)
s2.A[1,1] = s2.alpha0*100
s2.A[2,2] = s2.alpha1*100
s2.A[1,2] = rho.a*sqrt(s2.A[1,1])*sqrt(s2.A[2,2])
s2.A[2,1] = s2.A[1,2]

```

```

rho.a ~ dnorm(mu_rho.a, tau_rho.a)T(-1, 1)
mu_rho.a ~ dunif(-1, 1)
tau_rho.a ~ dgamma(1.5, 10E-4)
beta1 ~ dnorm(m.beta1, (s2.beta1*100)^(-1))T(0, 1)

# prior for std skewed-$t$
theta ~ dnorm(m.theta, (s2.theta*100)^(-1))
nu ~ dnorm(m.nu, (s2.nu*100)^(-1))T(2, )
}

```

References

- Ardia D and Hoogerheide LF (2009). Bayesian estimation of the GARCH(1, 1) model with Student- t innovations, *The R Journal*, **2**, 41–47.
- Baillie RT, Chung CF, and Tieslau MA (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics*, **11**, 23–40.
- Bollerslev T (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.
- Bollerslev T, Chou RY, and Kroner KF (1992). ARCH modeling in finance: a review of the theory and empirical evidence, *Journal of Econometrics*, **52**, 5–59.
- Corsi F, Mittnik S, Pigorsch C, and Pigorsch U (2008). The volatility of realized volatility, *Econometric Reviews*, **27**, 46–78.
- Czado C and Min A (2011). Bayesian inference for D-vines: estimation and model selection. In D Kurowicka and H Joe (Eds), *Dependence Modeling: Vine Copula Handbook* (pp. 249–264), World Scientific Publishing Co., Singapore.
- Dickey DA and Fuller WA (1979). Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.
- Geweke J and Porter-Hudak S (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis*, **4**, 221–238.
- Granger CW and Joyeux R (1980). An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, **1**, 15–29.
- Grassi S and de Magistris PS (2015). It's all about volatility of volatility: evidence from a two-factor stochastic volatility model, *Journal of Empirical Finance*, **30**, 62–78.
- Graves T, Gramacy RB, Franzke CLE, and Watkins NW (2015). Efficient Bayesian inference for ARFIMA processes, *Nonlinear Processes in Geophysics Discussions*, **2**, 573–618.
- Hosking JR (1981). Fractional differencing, *Biometrika*, **68**, 165–176.
- Kruschke JK (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (2nd ed), Academic Press, London.
- Kwiatkowski D, Phillips PCB, Schmidt P, and Shin Y (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159–178.
- Lambert P and Laurent S (2001). Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations, Discussion Paper, Retrieved September 1, 2017, from https://dial.uclouvain.be/pr/boreal/en/object/boreal%3A91014/datastream/PDF_01/view
- Maasoumi E and McAleer M (2008). Realized volatility and long memory: an overview, *Econometric Reviews*, **27**, 1–9.

- Ntzoufras I (2011). *Bayesian Modeling Using WinBUGS*, **698**, John Wiley & Sons, New York.
- Park SK (2016). *Value at risk forecasting for volatility index* (Master's thesis), Ewha Womans University, Seoul.
- Spiegelhalter DJ, Best NG, Carlin BP, and Van Der Linde A (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 583–639.
- Wang R, Kirby C, and Clark S (2013). Volatility of volatility, expected stock return and variance risk premium, In *26th Australasian Finance and Banking Conference 2013*.

Received July 6, 2017; Revised August 16, 2017; Accepted August 17, 2017