ELSEVIER

Journal of Econometrics 95 (2000) 57–69

JOURNAL OF Econometrics

www.elsevier.nl/locate/econbase

Bayesian analysis of ARMA–GARCH models: A Markov chain sampling approach

Teruo Nakatsuma*

Institute of Economic Research, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo, 186-8603, Japan

Received 1 August 1996; received in revised form 1 February 1999; accepted 1 April 1999

Abstract

We develop a Markov chain Monte Carlo method for a linear regression model with an ARMA(p, q)-GARCH(r, s) error. To generate a Monte Carlo sample from the joint posterior distribution, we employ a Markov chain sampling with the Metropolis–Hastings algorithm. As illustration, we estimate an ARMA–GARCH model of simulated time series data. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C11; C22

Keywords: ARMA process; Bayesian inference; GARCH; Markov chain Monte Carlo; Metropolis–Hastings algorithm

1. Introduction

In this paper, we propose a new Markov chain Monte Carlo (MCMC) method for Bayesian estimation and inference of the ARCH/GARCH model. Autoregressive conditional heteroskedasticity (ARCH) by Engle (1982) and generalized ARCH (GARCH) by Bollerslev (1986) have been extensively studied and applied in many fields of economics, especially in financial economics. Our

* Tel.: 042-580-8367; fax: 042-580-8333.
E-mail address: cr00387@srv.cc.hit-u.ac.jp (T. Nakatsuma)

PII: S0304-4076(99)00029-9

new MCMC method is applicable to estimating not only a simple ARCH/GARCH model but also a linear regression model with the ARMA–GARCH error, which we shall call the ARMA–GARCH model.

The MCMC method is one of Monte Carlo integration methods.[1] In this method, we generate samples of parameters in the model from their joint posterior distribution by Markov chain sampling, and evaluate complicated multiple integrals, which are necessary for Bayesian inference, by Monte Carlo integration. The Metropolis–Hastings (MH) algorithm is one of the Markov chain sampling schemes widely used among researchers. The MH algorithm, developed by Metropolis et al. (1953) and Hastings (1970), is a technique to generate random numbers from a probability distribution. This algorithm is quite useful especially when we cannot generate samples of the parameters directly from their joint posterior distribution. This is the case for the ARMA–GARCH model.

To develop a new MCMC method for the ARMA–GARCH model, we combine and modify Markov chain sampling schemes developed by Chib and Greenberg (1994) and Müller and Pole (1995). Chib and Greenberg (1994) designed a Markov chain sampling scheme for a linear regression model with an ARMA(p, q) error in which the disturbance term of the ARMA process was i.i.d. normal. Müller and Pole (1995) developed an MCMC procedure for a linear regression model with a GARCH error in which the error term of the regression model had no serial correlation. It seems natural to develop a new MCMC procedure for the ARMA–GARCH model by combining these two methods.

Organization of this paper is as follows. In Section 2, we explain our new MCMC method for the ARMA–GARCH model. In Section 3, we estimate ARMA–GARCH models with simulated data by our MCMC method as illustration. In Section 4, concluding remarks of this paper are given.

2. MCMC method for ARMA–GARCH models

2.1. ARMA–GARCH model and Bayesian inference

We consider the following linear regression model with an ARMA(p, q)–GARCH(r, s) error, or simply an ARMA–GARCH model:

$$y_t = x_t\gamma + u_t, \quad (t = 1, \ldots, n), \tag{1}$$

[1] Importance sampling, another popular Monte Carlo integration method, was also applied to ARCH and GARCH models by a few researchers (Geweke, 1989a, b; Kleibergen and Van Dijk, 1993, among others).

$$u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad \varepsilon_t | \mathscr{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2),$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{r} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \tag{2}$$

where y_t is a scalar of the dependent variable; x_t is a $1 \times k$ vector of the independent variables; γ is a $k \times 1$ vector of the regression coefficients; $\mathscr{F}_{t-1}$ is a σ-field generated by $\{y_{t-1}, y_{t-2}, \dots\}$; ϕ_js are the coefficients of the AR(p) process, θ_js are the coefficients of the MA(q) process; α_js and β_js are the coefficients of the GARCH(r, s) process. Let $\gamma = [\gamma_1, \dots, \gamma_k]'$, $\phi = [\phi_1, \dots, \phi_p]'$, $\theta = [\theta_1, \dots, \theta_q]'$, $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_r]'$, $\beta = [\beta_1, \dots, \beta_s]'$, $Y = [y_1, \dots, y_n]'$, $X = [x'_1, \dots, x'_n]'$, $\Phi(L) = \sum_{j=1}^{p} \phi_j L^j$, $\Theta(L) = \sum_{j=1}^{q} \theta_j L^j$, $A(L) = \sum_{j=1}^{r} \alpha_j L^j$, and $B(L) = \sum_{j=1}^{s} \beta_j L^j$ where L is the lag operator.

In practice, we often impose several constraints on parameters in the ARMA–GARCH model:

C1: all roots of $1 - \Phi(L) = 0$ are outside the unit circle,

C2: all roots of $1 + \Theta(L) = 0$ are outside the unit circle,

C3: $\alpha_j > 0$ for $j = 0, \dots, r$,

C4: $\beta_j > 0$ for $j = 1, \dots, s$.

C1 and C2 are related to stationarity and invertibility of the ARMA process. C3 and C4 are imposed to guarantee that the conditional variance σ_t^2 is always positive. In many previous studies of GARCH models, the following constraint:

C5: $A(1) + B(1) < 1$,

is imposed for the finiteness of the unconditional variance of ε_t. Since one of the objects in Bayesian analysis of the ARMA–GARCH model is to test whether the constraint C5 is true, we will not put C5 on the GARCH coefficients.

To perform Bayesian analyses of the ARMA–GARCH model, we construct the posterior density function of the model:

$$p(\delta | Y, X) = \frac{\ell(Y|X, \delta) p(\delta)}{\int \ell(Y|X, \delta) p(\delta) \, \mathrm{d}\delta}, \tag{3}$$

where δ is the set of all parameters in the ARMA–GARCH model, $\ell(Y|X, \delta)$ is the likelihood function, and $p(\delta)$ is the prior. The likelihood function of the ARMA–GARCH model is

$$\ell(Y|X, \delta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\hat{\varepsilon}_t^2}{2\sigma_t^2} \right], \tag{4}$$

where

$$\hat{\varepsilon}_t = \begin{cases} \varepsilon_0, & (t = 0) \\ y_t - x_t\gamma - \sum_{j=1}^{p} \phi_j (y_{t-j} - x_{t-j}\gamma) - \sum_{j=1}^{q} \theta_j \hat{\varepsilon}_{t-j}, & (t = 1, \dots, n) \end{cases} \tag{5}$$

and we assume $y_0 = \varepsilon_0$, $y_t = 0$ for $t < 0$, and $x_t = 0$ for $t \leqslant 0$. We treat the pre-sample error ε_0 as a parameter in the ARMA–GARCH model.

As the prior, we use the following proper prior:

$$\begin{aligned} p(\varepsilon_0, \gamma, \phi, \theta, \alpha, \beta) = {} & \mathrm{N}(\mu_{\varepsilon_0}, \Sigma_{\varepsilon_0}) \times \mathrm{N}(\mu_\gamma, \Sigma_\gamma) \\ & \times \mathrm{N}(\mu_\phi, \Sigma_\phi) I_{C1}(\phi) \times \mathrm{N}(\mu_\theta, \Sigma_\theta) I_{C2}(\theta) \\ & \times \mathrm{N}(\mu_\alpha, \Sigma_\alpha) I_{C3}(\alpha) \times \mathrm{N}(\mu_\beta, \Sigma_\beta) I_{C4}(\beta), \end{aligned} \tag{6}$$

where $I_{Cj}(\cdot)(j = 1, 2, 3, 4)$ is the indicator function which takes unity if the constraint holds; otherwise zero, and $\mathrm{N}(\cdot) I_{Cj}(\cdot)$ represents that the normal distribution $\mathrm{N}(\cdot)$ is truncated at the boundary of the support of the indicator function $I_{Cj}(\cdot)$.

In Bayesian inference, we need to evaluate the expectation of a function of parameters:

$$\mathrm{E}[f(\delta)] = \int f(\delta) p(\delta | Y, X) \, \mathrm{d}\delta. \tag{7}$$

The functional form of $f(\delta_j)$ depends on what kind of inference we conduct. For instance, when we estimate the posterior probability of $A(1) + B(1) < 1$, $f(\delta)$ is the indicator function $I(A(1) + B(1) < 1)$.

In the ARMA–GARCH model (1) and (2), however, it is difficult to evaluate the multiple integral in (7) analytically. We need to employ a numerical integration method. A Monte Carlo integration is one of widely used techniques for numerical integration. Let $\{\delta^{(1)}, \dots, \delta^{(m)}\}$ be samples generated from the posterior distribution $p(\delta | Y, X)$. Then (7) is approximated by

$$\mathrm{E}[f(\delta)] \approx \frac{1}{m} \sum_{i=1}^{m} f(\delta^{(i)}) \tag{8}$$

for enough large m.

To apply the Monte Carlo method, we need to generate samples $\{\delta^{(1)}, \dots, \delta^{(m)}\}$ from the posterior distribution. Since we cannot generate them directly, we use the Metropolis–Hastings (MH) algorithm, which is one of Markov chain sampling procedures such as the Gibbs sampler. In the MH algorithm we generate a value $\hat{\delta}$ from the proposal distribution $g(\delta)$ and accept the proposal value with probability:

$$\lambda(\delta, \hat{\delta}) = \min\left\{ \frac{p(\hat{\delta} | Y, X)/g(\hat{\delta})}{p(\delta | Y, X)/g(\delta)}, 1 \right\}. \tag{9}$$

See Tierney (1994) and Chib and Greenberg (1995) among others for further explanations and discussions about the MH algorithm as well as the MCMC methods.

2.2. MCMC procedure

We explain the outline of our new MCMC procedure for the ARMA–GARCH model. More details on the MCMC procedure are given in the appendix of this paper.

To construct a MCMC procedure for the ARMA–GARCH model, we divide the parameters into two groups. Let $\delta_1 = (\varepsilon_0, \gamma, \phi, \theta)$ be the first group and $\delta_2 = (\alpha, \beta)$ be the second. For each group of the parameters, we use different proposal distributions.

The proposal distributions for the first group δ_1 are based on the original ARMA–GARCH model:

$$y_t = x_t\gamma + \sum_{j=1}^{p} \phi_j(y_{t-j} - x_{t-j}\gamma) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma_t^2), \tag{10}$$

but we assume that the conditional variances $\{\sigma_t^2\}_{t=1}^n$ are fixed and known. Using (10), we can generate δ_1 from their proposal distributions by the MCMC procedure by Chib and Greenberg (1994) with some modifications.

The proposal distributions for the second group δ_2 are based on an approximated GARCH model:

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j}, \quad w_t \sim \mathrm{N}(0, 2\sigma_t^4), \tag{11}$$

where $l = \max\{r, s\}$, $\alpha_j = 0$ for $j > r$, and $\beta_j = 0$ for $j > s$. Model (11) is derived by using the well-known property of the GARCH model. As shown in Bollerslev (1986), the GARCH(r, s) model (2) is expressed as an ARMA(l, s) process of $\{\varepsilon_t^2\}_{t=1}^n$:

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + \tilde{w}_t - \sum_{j=1}^{s} \beta_j \tilde{w}_{t-j}, \tag{12}$$

where $\tilde{w}_t = \varepsilon_t^2 - \sigma_t^2$. Since $\tilde{w}_t = (\varepsilon_t^2/\sigma_t^2 - 1)\sigma_t^2 = (\chi^2(1) - 1)\sigma_t^2$, the conditional mean of $\tilde{w}_t$ is $\mathrm{E}(\tilde{w}_t|\mathscr{F}_{t-1}) = 0$, and the conditional variance is $\mathrm{Var}(\tilde{w}_t|\mathscr{F}_{t-1}) = 2\sigma_t^4$. Replacing $\tilde{w}_t$ with $w_t \sim \mathrm{N}(0, 2\sigma_t^4)$[2], we have (11). Given $\{\sigma_t^2\}_{t=1}^n$ and $\{\varepsilon_t^2\}_{t=1}^n$, we generate δ_2 from their proposal distributions by the MCMC procedure similar to Chib and Greenberg's method.

The outline of our MCMC procedure is as follows:

(a) generate $\delta_1 = (\varepsilon_0, \gamma, \phi, \theta)$ from the proposal distribution based on (10) given δ_2,

[2] You may suspect that this normal approximation is quite poor, but in our experience this approach works fine.

(b) generate $\delta_2 = (\alpha, \beta)$ from the proposal distribution based on (11) given δ_1,
(c) apply the MH algorithm after each parameter is generated in (a) or (b),
(d) repeat (a)–(c) until the sequences become stable.

In this procedure, we update $\{\varepsilon_t^2\}_{t=1}^n$ and $\{\sigma_t^2\}_{t=1}^n$ at every time after corresponding parameters are updated. The full description of our new procedure is given in the appendix of this paper.

3. An example of the MCMC estimation

In this section, we demonstrate how our new MCMC method is used in application. We consider the following regression model:

$$\begin{aligned}
y_t &= \gamma_1 + \gamma_2 x_t + u_t, \quad (t = 1, \dots, n)\\
u_t &= \phi_1 u_{t-j} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4},\\
&\quad \varepsilon_t | \mathscr{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2),\\
\sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \alpha_3 \varepsilon_{t-3}^2 + \alpha_4 \varepsilon_{t-4}^2 + \beta_1 \sigma_{t-1}^2\\
&\quad + \beta_2 \sigma_{t-2}^2,
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
&(\gamma_1, \gamma_2) = (1.0, 1.0),\\
&\phi_1 = 0.9,\\
&(\theta_1, \theta_2, \theta_3, \theta_4) = (-0.48, 0.36, -0.24, 0.12),\\
&(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.001, 0.24, 0.18, 0.12, 0.06),\\
&(\beta_1, \beta_2) = (0.2, 0.1),
\end{aligned}$$

x_t is drawn from the uniform distribution between -0.5 and 0.5, and we set $n = 1000$.

We estimate (13) by our new MCMC method and the maximum-likelihood estimation (MLE) for comparison. The maximum-likelihood estimates of the parameters are computed with the simulated annealing algorithm by Goffe (1996). For this data set, the covariance estimator $n^{-1}A_0^{-1}B_0A_0^{-1}$, where

$$A_0 = -\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial^2 \ln \ell_t(\delta)}{\partial\delta\partial\delta'}\right) \quad \text{and} \quad B_0 = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial \ln \ell_t(\delta)}{\partial\delta}\frac{\partial \ln \ell_t(\delta)}{\partial\delta'}\right),$$

is not positive definite at the maximum-likelihood estimates of δ because A_0 is not. Therefore, we use $n^{-1}B_0^{-1}$ as the covariance estimator.

The MCMC procedure is performed as follows. We generate 60,000 runs by the Markov chain sampling in a single-sample path. The number of runs to be discarded as the initial burn-in is determined by the convergence diagnostic (CD) in Geweke (1992). Let m_0 denote the number of runs to be discarded, and m the number of runs to be retained. Let $\bar{\delta}_1$ denote the sample mean of δ for the first m_1 runs in the sample path with m runs, and $\bar{\delta}_2$ for the last m_2 runs. As Geweke (1992) suggested, We set $m_1 = 0.1$ m and $m_2 = 0.5$ m. The test statistic for CD is given as

$$(\bar{\delta}_1 - \bar{\delta}_2)/[\hat{S}_1^{\delta}(0)/m_1 + \hat{S}_2^{\delta}(0)/m_2]^{1/2}, \tag{14}$$

where $\hat{S}_i^{\delta}(\cdot)$ is the spectrum density estimate for m_i runs, and (14) asymptotically follows the standard normal distribution. By this diagnostic, we choose $m_0 = 6000$ and $m = 54{,}000$.[3] Estimation results by the MLE and the MCMC are shown in Table 1.

In Table 1, the MLE and the MCMC produce comparable estimates for γ, ϕ, and θ. All t-ratios of these parameters in the MLE are greater than 2.0. For α and β, however, the t-ratios are less than 2.0 except for α_1. In the MCMC, the 95% interval does not include zero for γ, ϕ, and θ (this fact is guaranteed for α and β by assumption), and the true value is within the 95% interval for all parameters.

We also estimate the posterior probability of $A(1) + B(1) < 1$, the condition for the finite unconditional variance in the GARCH(p, q) process (2). This probability is estimated by (8) with $f(\delta) = I(A(1) + B(1) < 1)$. The estimated posterior probability is shown in the bottom of Table 1. The value, 0.917, is reasonable because $A(1) + B(1) = 0.24 + 0.18 + 0.12 + 0.06 + 0.2 + 0.1 = 0.9 < 1$ and the GARCH process in (13) has the finite unconditional variance.

4. Concluding remarks

In this paper, we derived a Markov chain Monte Carlo method with Metropolis–Hastings algorithm for Bayesian inference of a linear regression model with an ARMA–GARCH error,or the ARMA–GARCH model. Our new MCMC method allows the error term not only to be an ARMA(p, q) process, but also to have the GARCH variance. This flexibility is one of the major advantages of our new method. Moreover, our method can deal with the pre-sample values of the error term and conditional variance of the ARMA–GARCH model.

We also applied the new MCMC method to estimate an ARMA–GARCH model of simulated time series data, and demonstrated how to conduct the Bayesian inference with samples of the parameters obtained by the Markov

[3] We also tried to use every third, fifth or tenth point, instead of all points, in the sample path to avoid strong serial correlation. The changes in estimation results, however, were negligible.

Table 1
Estimation results

	True value		Posterior statistics			$\hat{\rho}$[d]
		MLE	Mean	s.d.	Median	CD[e]
γ_1	1.0	0.982	0.982	0.0131	0.982	0.143
		(0.0128)[a]	(7.35×10^5)[b]		[0.957,1.008][c]	− 0.551
γ_2	1.0	1.007	1.007	0.00564	1.007	0.168
		(0.00550)	(3.39×10^5)		[0.996,1.018]	− 0.435
ϕ_1	0.9	0.879	0.882	0.0184	0.883	0.706
		(0.0210)	(0.000366)		[0.844,0.916]	0.209
θ_1	− 0.48	− 0.433	− 0.437	0.0344	− 0.436	0.908
		(0.0382)	(0.000757)		[− 0.502, − 0.369]	1.43
θ_2	0.36	0.306	0.298	0.0357	0.298	0.903
		(0.0362)	(0.000774)		[0.230,0.368]	− 0.756
θ_3	− 0.24	− 0.228	− 0.228	0.0377	− 0.230	0.911
		(0.0362)	(0.000887)		[− 0.301, − 0.153]	− 0.0146
θ_4	0.12	0.161	0.159	0.0302	0.159	0.893
		(0.0337)	(0.000620)		[0.0990,0.218]	0.600
ε_0	—	0.168	0.0935	0.259	0.151	0.373
		(0.168)	(0.00173)		[− 0.424,0.554]	0.0750
α_0	0.001	0.000774	0.000829	0.000200	0.000815	0.691
		(0.000705)	(5.42×10^6)		[0.000473,0.00126]	− 0.502
α_1	0.24	0.223	0.228	0.0509	0.228	0.406
		(0.0562)	(0.000406)		[0.130,0.329]	0.507
α_2	0.18	0.114	0.144	0.0586	0.144	0.546
		(0.246)	(0.000997)		[0.0295,0.260]	− 0.660
α_3	0.12	0.139	0.134	0.0603	0.132	0.604
		(0.112)	(0.00140)		[0.0225,0.260]	0.276
α_4	0.06	0.0613	0.0688	0.0446	0.0627	0.566
		(0.157)	(0.000939)		[0.00395,0.168]	− 0.746
β_1	0.2	0.373	0.236	0.131	0.233	0.805
		(1.08)	(0.00369)		[0.0158,0.491]	− 0.100
β_2	0.1	0.0238	0.123	0.0855	0.108	0.624
		(0.559)	(0.000927)		[0.00558,0.314]	1.17
$P\{A(1) + B(1) < 1\}$			0.917	0.275	—	0.249
			(0.00165)			0.700

[a]the standard error of the MLE.
[b]the numerical standard error of the posterior mean.
[c]95% interval, $[Q_{2.5\%}, Q_{97.5\%}]$.
[d]the first-order autocorrelation in a sample path.
[e]convergence diagnostic test statistic (14).

chain sampling. We estimated the posterior statistics of the parameters and compare them to the MLE. We also estimated the posterior probability of the finiteness of the unconditional variance in the GARCH(r, s) process, which is difficult to test in the classical approach.

Appendix A

In this Appendix, we explain the proposal distributions of parameters in the ARMA–GARCH model for the MCMC procedure. The parameters to be generated are (i) pre-sample error ε_0; (ii) regression coefficients γ; (iii) AR coefficients ϕ; (iv) MA coefficients θ; (v) ARCH coefficients α; (vi) GARCH coefficients β. We also discuss how to deal with the pre-sample error of the approximated GARCH model (11), w_0.

A.1. ε_0: *Pre-sample error*

Given the pre-sample error ε_0, the ARMA–GARCH model is rewritten as

$$y_t = x_t\gamma + \sum_{j=1}^{p} \phi_j(y_{t-j} - x_{t-j}\gamma) + \varepsilon_t + \sum_{j=1}^{q} \theta_j\varepsilon_{t-j} + (\phi_t + \theta_t)\varepsilon_0, \tag{A.1}$$

where $y_t = \varepsilon_t = 0$ for $t < 0$, $x_t = 0$ for $t < 0$, $\phi_t = 0$ for $t > p$, and $\theta_t = 0$ for $t > q$. In (A.1), y_t does not depend on ε_0 for $t > \max\{p, q\}$. Chib and Greenberg (1994) derived a similar expression for a regression model with an ARMA(p, q) error.

The likelihood function of the ARMA–GARCH model is rewritten as

$$f(Y|X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{(y_t^\dagger - x_t^\dagger\varepsilon_0)^2}{2\sigma_t^2} \right], \tag{A.2}$$

where $y_t^\dagger$ and $x_t^\dagger$ are computed by

$$\begin{aligned} y_t^\dagger &= y_t - x_t\gamma - \sum_{j=1}^{p} \phi_j(y_{t-j} - x_{t-j}\gamma) - \sum_{j=1}^{q} \theta_j y_{t-j}^\dagger, \\ x_t^\dagger &= (\phi_t + \theta_t) - \sum_{j=1}^{q} \theta_j x_{t-j}^\dagger, \end{aligned} \tag{A.3}$$

with $y_t = y_t^\dagger = 0$ for $t \leqslant 0$, and $x_t = x_t^\dagger = 0$ for $t \leqslant 0$. Obviously, $\varepsilon_t = y_t^\dagger - x_t^\dagger\varepsilon_0$. We have the proposal distribution of ε_0:

$$\varepsilon_0|Y, X, \Sigma \sim \mathrm{N}(\hat{\mu}_{\varepsilon_0}, \hat{\sigma}_{\varepsilon_0}), \tag{A.4}$$

where $\hat{\mu}_{\varepsilon_0} = \hat{\Sigma}_{\varepsilon_0}(\sum_{t=1}^{n} x_{\varepsilon t}^2/\sigma_t^2 + \mu_{\varepsilon_0}/\sigma_{\varepsilon_0}^2)$ and $\hat{\Sigma}_{\varepsilon_0} = (\sum_{t=1}^{n} x_{\varepsilon t}y_{\varepsilon t}/\sigma_t^2 + \sigma_{\varepsilon_0}^{-2})^{-1}$.

A.2. γ: *Regression coefficients*

The likelihood function of the ARMA–GARCH model is rewritten as

$$f(Y|X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{(y_t^* - x_t^*\gamma)^2}{2\sigma_t^2} \right], \tag{A.5}$$

where y_t^* and x_t^* are calculated by the following transformation:

$$\begin{aligned} y_t^* &= y_t - \sum_{j=1}^{p} \phi_j y_{t-j} - \sum_{j=1}^{q} \theta_j y_{t-j}^*, \\ x_t^* &= x_t - \sum_{j=1}^{p} \phi_j x_{t-j} - \sum_{j=1}^{q} \theta_j x_{t-j}^*, \end{aligned} \tag{A.6}$$

where $y_0^* = \varepsilon_0$, $y_t = y_t^* = 0$ for $t < 0$, and $x_t = x_t^* = 0$ for $t \leqslant 0$. This is a modified version of the transformation derived by Chib and Greenberg (1994). It is straightforward to show $\varepsilon_t = y_t^* - x_t^* \gamma$. Let $Y_\gamma = [y_1^*, \dots, y_n^*]'$ and $X_\gamma = [x_1^{*'}, \dots, x_n^{*'}]'$. We have the following proposal distribution of γ:

$$\gamma | Y, X, \Sigma, \delta_{-\gamma} \sim \mathrm{N}(\hat{\mu}_\gamma, \hat{\Sigma}_\gamma), \tag{A.7}$$

where $\hat{\mu}_\gamma = \hat{\Sigma}_\gamma (X_\gamma' \Sigma^{-1} Y_\gamma + \Sigma_\gamma^{-1} \mu_\gamma)$ and $\hat{\Sigma}_\gamma = (X_\gamma' \Sigma^{-1} X_\gamma + \Sigma_\gamma^{-1})^{-1}$.

A.3. ϕ: AR coefficients

The likelihood function of the ARMA–GARCH model is also written as

$$f(Y|X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{(\tilde{y}_t - \tilde{x}_t \phi)^2}{2\sigma_t^2} \right], \tag{A.8}$$

where $\tilde{y}_t$ and $\tilde{x}_t$ are calculated by the following transformation:

$$\tilde{y}_t = y_t - x_t \gamma - \sum_{j=1}^{q} \theta_j \tilde{y}_{t-j}, \quad \tilde{x}_t = [\tilde{y}_{t-1}, \dots, \tilde{y}_{t-p}], \tag{A.9}$$

where $\tilde{y}_0 = \varepsilon_0$, $y_t = \tilde{y}_t = 0$ for $t < 0$. This is also a modified version of the transformation derived by Chib and Greenberg (1994). It is also straightforward to show $\varepsilon_t = \tilde{y}_t - \tilde{x}_t \phi$. Let $Y_\phi = [\tilde{y}_1, \dots, \tilde{y}_n]'$ and $X_\phi = [\tilde{x}'_1, \dots, \tilde{x}'_n]'$. We have the proposal distribution of ϕ:

$$\phi | Y, X, \Sigma, \delta_{-\phi} \sim \mathrm{N}(\hat{\mu}_\phi, \hat{\Sigma}_\phi) I_{C1}(\phi), \tag{A.10}$$

where $\hat{\mu}_\phi = \hat{\Sigma}_\phi (X_\phi' \Sigma^{-1} Y_\phi + \Sigma_\phi^{-1} \mu_\phi)$ and $\hat{\Sigma}_\phi = (X_\phi' \Sigma^{-1} X_\phi + \Sigma_\phi^{-1})^{-1}$.

A.4. θ: MA coefficients

Generation of θ is a little more complicated since the error term u_t is a non-linear function of θ. To deal with this complexity, Chib and Greenberg (1994) proposed to linearize ε_t by the first-order Taylor expansion

$$\varepsilon_t(\theta) \approx \varepsilon_t(\theta^*) + \psi_t(\theta - \theta^*), \tag{A.11}$$

where $\varepsilon_t(\theta^*) = y_t^*(\theta^*) - x_t^*(\theta^*)$ and $\psi_t = [\psi_{1t}, \dots, \psi_{qt}]$ is the first-order derivative of $\varepsilon_t(\theta)$ evaluated at θ^* given by the following recursion:

$$\psi_{it} = -\varepsilon_{t-i}(\theta^*) - \sum_{j=1}^{q} \theta_j^* \psi_{it-j} \quad (i = 1, \dots, q), \tag{A.12}$$

where $\psi_{it} = 0$ for $t \leqslant 0$. The choice of θ^* is crucial to obtain a suitable approximation. Chib and Greenberg (1994) chose the non-linear least-squares estimate of θ given the other parameters as θ^*,

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \{\varepsilon_t(\theta)\}^2. \tag{A.13}$$

However, the error term in the ARMA–GARCH model is heteroskedastic while Chib and Greenberg applied their approximation to the homoskedastic-error ARMA model. Thus, instead of (A.13), we use the following weighted non-linear least-squares estimate of θ:

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{n} \{\varepsilon_t(\theta)\}^2/\sigma_t^2, \tag{A.14}$$

to approximate the likelihood function for model (10).

Then we have an approximated likelihood function for model (10):

$$f(Y|X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{\{\varepsilon_t(\theta^*) + \psi_t(\theta - \theta^*)^2\}^2}{2\sigma_t^2} \right]. \tag{A.15}$$

Let $Y_\theta = [\psi_1\theta^* - \varepsilon_1(\theta^*), \dots, \psi_n\theta^* - \varepsilon_n(\theta^*)]'$ and $X_\theta = [\psi_1', \dots, \psi_n']'$. Using the approximated likelihood function, we have the following proposal distribution of θ:

$$\theta | Y, X, \Sigma, \delta_{-\theta} \sim \mathrm{N}(\hat{\mu}_\theta, \hat{\Sigma}_\theta) I_{C2}(\theta), \tag{A.16}$$

where $\hat{\mu}_\theta = \hat{\Sigma}_\theta(X_\theta'\Sigma^{-1}Y_\theta + \Sigma_\theta^{-1}\mu_\theta)$ and $\hat{\Sigma}_\theta = (X_\theta'\Sigma^{-1}X_\theta + \Sigma_\theta^{-1})^{-1}$.

A.5. w_0: Pre-sample error in the approximated GARCH model

Before we explain the proposal distributions for α and β, we discuss how to deal with the pre-sample error in the approximated GARCH model (11), w_0. In (11), ε_t^2 is given as

$$\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{l} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + w_t - \sum_{j=1}^{s} \beta_j w_{t-j} + \alpha_t w_0, \tag{A.17}$$

where $\varepsilon_t^2 = w_t = 0$ for $t < 0$, $\alpha_t = 0$ for $t > r$, and $\beta_t = 0$ for $t > s$. Note that ε_t^2 does not depend on w_0 for $t > r$.

Unlike ε_0, we do not need to generate w_0. Suppose that the initial variance σ_0^2 is equal to α_0. Since $w_t = \varepsilon_t^2 - \sigma_t^2$, we have

$$w_0 = \varepsilon_0^2 - \alpha_0. \tag{A.18}$$

Hence, w_0 is automatically determined by (A.18) once we obtain ε_0 and α_0.

A.6. α: ARCH coefficients

Generation of α is similar to ϕ. The likelihood function of the approximated GARCH model (11) is rewritten as

$$f(\varepsilon^2|Y, X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi(2\sigma_t^4)}} \exp\left[-\frac{(\bar{\varepsilon}_t^2 - \zeta_t\alpha)^2}{2(2\sigma_t^4)} \right], \tag{A.19}$$

where $\varepsilon^2 = [\varepsilon_1^2, \ldots, \varepsilon_n^2]'$, $\bar{\varepsilon}_t^2 = \tilde{\varepsilon}_t^2 - \sum_{j=1}^{s} \beta_j \tilde{\varepsilon}_{t-j}^2$, and $\zeta_t = [\tilde{\iota}_t, \tilde{\varepsilon}_{t-1}^2, \ldots, \tilde{\varepsilon}_{t-r}^2]$. $\tilde{\varepsilon}_t^2$ and $\tilde{\iota}_t$ are obtained by transformation (A.9) with $\theta_j \to -\beta_j$ and $y_t - x_t\gamma \to \varepsilon_t^2$ for $\tilde{\varepsilon}_t^2$, or $y_t - x_t\gamma \to 1$ for $\tilde{\iota}_t$. Obviously, $w_t = \bar{\varepsilon}_t^2 - \zeta_t\alpha$. Let $Y_\alpha = [\bar{\varepsilon}_1^2, \ldots, \bar{\varepsilon}_n^2]'$ and $X_\alpha = [\zeta_1', \ldots, \zeta_n']'$. Then we have the following proposal distribution of α:

$$\alpha|Y, X, \Sigma, \delta_{-\alpha} \sim \mathrm{N}(\hat{\mu}_\alpha, \hat{\Sigma}_\alpha) I_{C3}(\alpha), \tag{A.20}$$

where $\hat{\mu}_\alpha = \hat{\Sigma}_\alpha(X_\alpha' \Lambda^{-1} Y_\alpha + \Sigma_\alpha^{-1}\mu_\alpha)$, $\hat{\Sigma}_\alpha = (X_\alpha' \Lambda^{-1} X_\alpha + \Sigma_\alpha^{-1})^{-1}$, and $\Lambda = \mathrm{diag}\{2\sigma_1^4, \ldots, 2\sigma_n^4\}$.

A.7. β: GARCH coefficients

Generation of β is similar to θ. We approximate the likelihood function of (11) as

$$f(\varepsilon^2|Y, X, \Sigma, \delta_1, \delta_2) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi(2\sigma_t^4)}} \exp\left[-\frac{\{w_t(\beta^*) - \xi_t(\beta - \beta^*)\}^2}{2(2\sigma_t^4)} \right], \tag{A.21}$$

where $\xi_t = [\xi_{1t}, \ldots, \xi_{st}]$ is the first-order derivative of $\sigma_t^2(\beta)$ evaluated at β^*, which is computed by the transformation (A.12) with $-\varepsilon_{t-i}(\theta^*) \to \sigma_{t-i}^2(\beta^*)$ and $\theta_j^* \to -\beta_j^*$, and β^* is the solution of (A.14) with $\theta \to \beta$, $\varepsilon_t(\theta) \to w_t(\beta)$, and $\sigma_t^2 \to 2\sigma_t^4$. Let $Y_\beta = [w_1(\beta^*) + \xi_1\beta^*, \ldots, w_n(\beta^*) + \xi_n\beta^*]'$ and $X_\beta = [\xi_1', \ldots, \xi_n']'$. Using the approximated likelihood function (A.21), we have the following proposal distribution of β:

$$\beta|Y, X, \Sigma, \delta_{-\beta} \sim \mathrm{N}(\hat{\mu}_\beta, \hat{\Sigma}_\beta) I_{C4}(\beta), \tag{A.22}$$

where $\hat{\mu}_\beta = \hat{\Sigma}_\beta(X_\beta' \Lambda^{-1} Y_\beta + \Sigma_\beta^{-1}\mu_\beta)$ and $\hat{\Sigma}_\beta = (X_\beta' \Lambda^{-1} X_\beta + \Sigma_\beta^{-1})^{-1}$.

References

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA (p, q) errors. Journal of Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. American Statistician 49, 327–335.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1008.

Geweke, J., 1989a. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Geweke, J., 1989b. Exact predictive densities for linear models with ARCH disturbances. Journal of Econometrics 40, 63–86.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 4. Oxford University Press, Oxford, pp. 169–193.

Goffe, W.L., 1996. SIMANN: a global optimization algorithm using simulated annealing. Studies in Nonlinear Dynamics and Econometrics 1, 169–176.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Kleibergen, F., Van Dijk, H.K., 1993. Non-stationarity in GARCH models: A Bayesian analysis. Journal of Applied Econometrics 8, S41–S61.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1092.

Müller, P., Pole, A., 1995. Monte Carlo posterior integration in GARCH models. Manuscript, Duke University.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Annals of Statistics 22, 1701–1762.