Full Bayesian Inference for GARCH and EGARCH Models

I. D. VRONTOS and P. DELLAPORTAS
Department of Statistics, Athens University of Economics and Business, 10434 Athens, Greece

D. N. POLITIS
Department of Mathematics, University of California at San Diego, La Jolla, CA 92093-0112

A full Bayesian analysis of GARCH and EGARCH models is proposed consisting of parameter estimation, model selection, and volatility prediction. The Bayesian paradigm is implemented via Markov-chain Monte Carlo methodologies. We provide implementation details and illustrations using the General Index of the Athens stock exchange.

KEY WORDS: Markov-chain Monte Carlo; Model averaging; Reversible jump; Volatility prediction.

Autoregressive conditional heteroscedasticity (ARCH) models (Engle 1982) and their extended versions have been proven to be very successful in modeling the volatility of financial time series. By postulating the time-varying volatility to be a function of the current information set, they are able to model the periods of relative tranquillity followed by bursts of extreme values often present in stock-market series. Despite both the attractiveness and the complexity of these models, the Bayesian community has not yet broadly utilized the recent advances in statistical computation based on Markov-chain Monte Carlo (MCMC) methods to effect a practical statistical analysis "package." This article considers the general problem of Bayesian inference, prediction, and model criticism of ARCH-type models. We demonstrate that adoption of the Bayesian framework can be advantageous on grounds of generality, accuracy, and flexibility. Moreover, the MCMC sampling-based approach provides an idealized way to extract any posterior summary of interest such as functions of parameters and in addition to construct predictive densities that take into account model uncertainty.

Two of the most useful ARCH parameterizations are the generalized ARCH (GARCH) model introduced by Bollerslev (1986) and the exponential GARCH (EGARCH) model suggested by Nelson (1991). Although the list of other proposed ARCH parameterizations is long (e.g., see Engle 1995), in this article we focus on GARCH and EGARCH models, without restricting the methodological potential of our suggested procedures. GARCH models are chosen because of their extended applicability to many financial data and EGARCH models because of both their good performance in a series of comparative studies (see Pagan and Schwert 1990; Engle and Ng 1993; Shephard 1996) and their complexity, which can be viewed as a challenge for our methodology. In ARCH-type models, the estimation of the parameters can be done by using maximum likelihood, quasi-maximum likelihood, or generalized method of moments; for details, see the review article by Bollerslev, Chou, and Kroner (1992).

The key steps in our proposed framework are as follows. First, within a certain GARCH/EGARCH model, we construct a Markov chain that has as a stationary distribution the posterior distribution of the model parameters. Simulation of this Markov chain provides, after some burn-in period and sufficient iterations, samples from the posterior distribution of interest; for details, see Smith and Roberts (1993) or Besag, Green, Higdon, and Mengersen (1995). We provide detailed guidelines on how to construct the required Markov chain using Metropolis–Hastings steps.

Second, for a given set of competing models, we propose modeling each GARCH or EGARCH model jointly and base our inference about the models on their posterior probabilities or Bayes factors. Thus, we avoid the usual approach, which considers the models separately and chooses the best model via significance tests. We believe that the joint estimation of parameters and model probabilities not only provides a probabilistically sound way to overcome the awkward model-selection problem in GARCH/EGARCH models but also introduces a new way to predict the future volatility via "model averaging." To obtain a sample of the joint posterior density of models and model parameters, we extend the MCMC strategy so that the sampler jumps between parameter subspaces of different dimensionality corresponding to different models. This idea is based on reversible-jump MCMC introduced by Green (1995). A Bayesian strategy similar to the one we propose for the parameter estimation of our models was described by Muller and Pole (in press). It is an MCMC strategy that handles successfully a GARCH model with covariates by using Metropolis–Hastings steps with a sophisticated choice of proposal distributions. A pioneer Bayesian implementation strategy for such models is due to Geweke (1989), who proposed a Monte Carlo strategy to derive the desired posterior summaries of interest. Finally, in a similar context, Jacquier, Polson, and Rossi (1994) proposed a Metropolis–Hastings algorithm to derive posterior distributions of the parameters of stochastic volatility models.

Journal of Business & Economic Statistics
April 2000, Vol. 18, No. 2

The remainder of the article is organized as follows. GARCH and EGARCH models are briefly introduced in Section 1. MCMC methodologies including detailed implementation guidelines for the recently proposed reversible-jump MCMC are presented in Section 2. In Section 3, we present a Monte Carlo simulation for GARCH(1,1) and EGARCH(1,1) models and compare the "true" volatilities with the predictive volatilities from classical and Bayesian techniques. In Section 4, we illustrate our proposed methodology in GARCH and EGARCH models and apply the reversible-jump MCMC method using data from the Athens stock exchange. Finally, we conclude in Section 5 with a brief discussion.

1. GARCH AND EGARCH MODELS

Let $\{\varepsilon_t\}$ be a real-valued discrete-time stochastic process, which is the error process. Assume that the mean equation is of the form $y_t = \varepsilon_t$, where y_t is the observation process. The GARCH(p, q) model (Bollerslev 1986) is given by the following two-stage formulation:

$$\varepsilon_t = z_t\sigma_t, \qquad t = 0, \ldots, T, \tag{1}$$

and

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \qquad t = 1, \ldots, T, \tag{2}$$

where z_t are iid with $E(z_t) = 0$ and $\operatorname{var}(z_t) = 1; \sigma_t^2$ is the (conditional) variance of the $\{\varepsilon_t\}$ process at time $t; p, q$ are integers with $p > 0, q \geq 0, \alpha_0 > 0$, and $\alpha_i \geq 0, i = 1, \ldots, p$; and $\beta_j \geq 0, j = 1, \ldots, q$. In (2) it is assumed that $\varepsilon_t = \sigma_t = 0$ for $t < 0$. These restrictions ensure a positive variance. Stationarity conditions impose that $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$; for details, see, Bollerslev (1986).

The preceding family of models is composed of generalizations of the early work of Engle (1982). They achieve more flexible lag structure than ARCH models and capture the volatility clustering phenomenon. Actually, GARCH models are more parsimonious in that, although typically a high-order ARCH is required to model a particular series, a GARCH(1,1), say, might do equally well; see, for example, Bollerslev (1986) or Shephard (1996). An invertible GARCH model is, in fact, an infinite-order ARCH model with restrictions on the coefficients.

Assuming that z_t are normally distributed, the parameter vector to be estimated in (1) and (2) is, for $q > 0, \boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma_0^2)$. The likelihood for a sample of $T + 1$ observations $\mathbf{y} = (y_1, \ldots, y_{T+1})$ can be written as

$$l_N(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-(T+1)/2} \prod_{t=0}^{T} \left\{ (\sigma_t^2)^{-1/2} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \right\},$$

where σ_t^2 is expressed via (2). Under the assumption of a Student-t distribution (Bollerslev 1987; Baillie and Bollerslev 1989) for the error process $\{\varepsilon_t\}$, the likelihood for a sample of $T + 1$ observations $\mathbf{y}$ can be written as

$$l_T(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=0}^{T} \left\{ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)[(n-2)\sigma_t^2]^{1/2}} \left(1 + \frac{\varepsilon_t^2}{(n-2)\sigma_t^2}\right)^{-(n+1)/2} \right\},$$

where $n > 2$ denotes the degrees of freedom of the Student-t distribution and the parameter vector to be estimated is $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma_0^2, n)$.

A modification of GARCH models is the EGARCH models introduced by Nelson (1991). They are defined by assuming that z_t in (1) follows a generalized error distribution with mean 0, variance 1, and parameter v so that, if $v = 2$, we obtain the normal distribution, and for $v < 2$ and $v > 2$, the distribution of z_t has thicker and thinner tails than the normal distribution, respectively. Moreover, the EGARCH formulation in the second stage of Equation (2) is given by

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{q} \beta_j \ln(\sigma_{t-j}^2) + \sum_{k=1}^{p} [\theta_k z_{t-k} + \gamma_k(|z_{t-k}| - E|z_{t-k}|)], \tag{3}$$

where $\sigma_t = z_t = 0$ for $t < 0$ and $E|z_{t-k}|$ is given, under the assumption of generalized error distribution for z_t, by

$$E|z_{t-k}| = \frac{\Gamma\left(\frac{2}{v}\right)}{\left[\Gamma\left(\frac{1}{v}\right)\Gamma\left(\frac{3}{v}\right)\right]^{1/2}}.$$

The preceding EGARCH(p, q) formulation has parameters $\boldsymbol{\theta} = (\alpha_0, \beta_1, \ldots, \beta_q, \theta_1, \ldots, \theta_p, \gamma_1, \ldots, \gamma_p, v, \sigma_0^2)$ and the likelihood function is given by

$$l(\mathbf{y}|\boldsymbol{\theta}) = \left[\frac{v}{\lambda 2^{1+(1/v)}\Gamma\left(\frac{1}{v}\right)}\right]^{T+1} \times \prod_{t=0}^{T} \left\{ (\sigma_t^2)^{-1/2} \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_t}{\sigma_t\lambda}\right|^v\right) \right\}. \tag{4}$$

Under the usual assumption that the past variance σ_0^2 is specified in advance, then the EGARCH(p, q) model has a parameter vector $\boldsymbol{\theta}^* = \boldsymbol{\theta} \setminus \{\sigma_0^2\}$ and the likelihood for a sample of $T + 1$ observations can be written as

$$l(\mathbf{y}|\boldsymbol{\theta}^*) = \left[\frac{v}{\lambda 2^{1+(1/v)}\Gamma\left(\frac{1}{v}\right)}\right]^{T-p+1} \times \prod_{t=p}^{T} \left\{ (\sigma_t^2)^{-1/2} \exp\left(-\frac{1}{2}\left|\frac{\varepsilon_t}{\sigma_t\lambda}\right|^v\right) \right\}.$$

2. BAYESIAN FORMULATION AND IMPLEMENTATION

2.1 Inferences for a Given Model

Bayesian inferences about the parameter vector $\boldsymbol{\theta} =$

$(\theta_1, \ldots, \theta_n)$ conditional on data $\mathbf{y}$ are made via the posterior density $\pi(\boldsymbol{\theta}|\mathbf{y})$. Using the Bayes theorem, this density takes the form $\pi(\boldsymbol{\theta}|\mathbf{y}) = cl(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ for some normalizing constant c, likelihood function $l(\mathbf{y}|\boldsymbol{\theta})$, and prior density $\pi(\boldsymbol{\theta})$.

For many realistic problems, evaluation of $\pi(\boldsymbol{\theta}|\mathbf{y})$ is analytically intractable, so numerical or asymptotic methods are necessary to obtain posterior summaries of interest; see Evans and Swartz (1995) for a recent review of possible avenues. In this article, we adopt the MCMC sampling strategies as our tool for this purpose. The idea is based on the construction of an irreducible and aperiodic Markov chain with realizations $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^t, \ldots$ in the parameter space, equilibrium distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$, and a transition probability $K(\boldsymbol{\theta}'', \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}''|\boldsymbol{\theta}^t = \boldsymbol{\theta}')$, where $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$ are the realized states at time t and $t+1$, respectively. Under appropriate regularity conditions, asymptotic results guarantee that, as $t \to \infty, \boldsymbol{\theta}^t$ tends in distribution to a random variable with density $\pi(\boldsymbol{\theta}|\mathbf{y})$, and the ergodic average of an integrable function of $\boldsymbol{\theta}$ is a consistent estimator of the (posterior) mean of the function.

There are many possible choices for the transition kernel K, each leading to different sampling schemes. A simple strategy is to use n independent Metropolis steps (Tierney 1994; Chib and Greenberg 1995) for all $\theta_i, i = 1, 2, \ldots, n$, and a usual approach is to adopt a random-walk chain with an increment normal density $\mathrm{N}(0, \sigma^2)$, where the variance σ^2 is appropriately chosen so that the convergence of the MCMC sampler is as fast as possible. Metropolis steps are probably not the best choice for constructing the required Markov chain, but if the full conditional posterior densities (i.e., the densities of each element or subvector of $\boldsymbol{\theta}$ given all the other elements) are not of known form, or if they do not have exploitable properties such as log-concavity of known maximum, Metropolis steps give the easiest black-box sampling strategy yielding the required realizations of $\pi(\boldsymbol{\theta}|\mathbf{y})$.

Despite the fact that we propose this simple sampling scheme, in which we update the elements of the parameter vector one at a time (single-component update), we emphasize in the application in Section 4 that other alternatives may drastically improve the efficiency of the algorithm. For example, our empirical data analysis indicated that simultaneous sampling of a carefully chosen subvector of $\boldsymbol{\theta}$, after a possible transformation to a subvector taking values on $(-\infty, \infty)$, may be extremely appealing. To do that, we update the highly correlated elements of the parameter vector simultaneously (simultaneous vector update), and take a sample from this subvector of $\boldsymbol{\theta}$ using multivariate Metropolis steps as follows. We first estimate the sample covariance matrix $\boldsymbol{\Sigma}$ related to this subvector from an initial exploratory run of the Markov chain. Then, the update from time t to time $t+1$ is achieved by using a multivariate normal proposal density $\mathrm{N}(\boldsymbol{\mu}^t, c\boldsymbol{\Sigma})$, with $\boldsymbol{\mu}^t$ denoting the vector at time t and c a constant to tune the acceptance rate.

Muller and Pole (in press) dealt with the correlation problem in estimating the parameters of a GARCH(1,1) model in another way. They chose their proposal distribution to mimic the true full conditional posterior distribution by using related results from regression problems. An extra cost is an additional rejection step, which is needed to maintain the joint posterior density as the stationary distribution of the simulated Markov chain. The authors reported problems in the preceding methodology when the proposal distribution is significantly thinner than the true full conditional, and they proposed a rather sophisticated periodic change of the shape of the proposal density.

2.2 Inferences Under Model Uncertainty

Assume that we have a countable set M of competing models for a given set of data $\mathbf{y}$. Let model $m \in M$ have a vector $\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m$ of unknown parameters, the dimension of which may vary from model to model. The posterior probability of model m is given by

$$\pi(m|\mathbf{y}) = \frac{\pi(m) \int_{\boldsymbol{\Theta}_m} \pi(\mathbf{y}|m, \boldsymbol{\theta}_m)\pi(\boldsymbol{\theta}_m|m)\, d\boldsymbol{\theta}_m}{\sum_{m \in M} \pi(m) \int_{\boldsymbol{\Theta}_m} \pi(\mathbf{y}|m, \boldsymbol{\theta}_m)\pi(\boldsymbol{\theta}_m|m)\, d\boldsymbol{\theta}_m}, \quad (5)$$

where $\pi(\mathbf{y}|m, \boldsymbol{\theta}_m)$ is the likelihood given the model m, the parameter vector $\boldsymbol{\theta}_m, \pi(m)$ is the prior probability for model m, and $\pi(\boldsymbol{\theta}_m|m)$ is the prior of the parameter vector $\boldsymbol{\theta}_m$ given the model m. Inference about the model-selection problem may be done using the Bayes factor (BF) of model m_i against model m_j given by

$$\begin{aligned} \mathrm{BF} &= \frac{\pi(m_i|\mathbf{y})}{\pi(m_j|\mathbf{y})} \frac{\pi(m_j)}{\pi(m_i)} \\ &= \frac{\int_{\boldsymbol{\Theta}_{m_i}} \pi(\mathbf{y}|m_i, \boldsymbol{\theta}_{m_i})\pi(\boldsymbol{\theta}_{m_i}|m_i)\, d\boldsymbol{\theta}_{m_i}}{\int_{\boldsymbol{\Theta}_{m_j}} \pi(\mathbf{y}|m_j, \boldsymbol{\theta}_{m_j})\pi(\boldsymbol{\theta}_{m_j}|m_j)\, d\boldsymbol{\theta}_{m_j}}. \end{aligned} \quad (6)$$

Kass and Raftery (1995) gave a series of arguments that make Bayes factors appealing when compared with other model-selection strategies such as the Akaike information criterion (AIC) or the Bayes information criterion (BIC). Bayes factors, however, require evaluation of the integrals in the numerator and denominator of (6), which are the marginal densities $\pi(\mathbf{y}|m_i)$ and $\pi(\mathbf{y}|m_j)$. These integrals are, in general, difficult to calculate; Kass and Raftery (1995) provided an extensive description and comparison of available numerical strategies.

Green (1995) introduced a reversible-jump MCMC strategy for generating from the joint posterior $\pi(m, \boldsymbol{\theta}_m|\mathbf{y})$, based on the standard Metropolis–Hastings approach. The reversible-jump MCMC was also applied by Richardson and Green (1997) for an analysis of univariate normal mixtures, Nobile and Green (in press) for factorial experiments using mixture modeling, and Dellaportas and Forster (1999) for analysis of contingency tables. During reversible-jump MCMC sampling, the constructed Markov chain moves within and between models so that the limiting proportion of visits to a given model is the required $\pi(m|\mathbf{y})$ in (5).

In general, suppose that the current state of the Markov chain at time t is $(m, \boldsymbol{\theta}_m)$, where $\boldsymbol{\theta}_m$ has dimension $d(\boldsymbol{\theta}_m)$ and a move is proposed at time $t+1$ to a new model m' with probability $j(m, m')$ and corresponding parameter vector $\boldsymbol{\theta}'_{m'}$. Then, a vector $\mathbf{u}$ is generated from a specified

proposal density $q(\mathbf{u}|\boldsymbol{\theta}_m, m, m')$, and we set $(\boldsymbol{\theta}'_{m'}, \mathbf{u}') = g_{m,m'}(\boldsymbol{\theta}_m, \mathbf{u})$ for a specified invertible function $g_{m,m'}$ such that $g_{m',m} = g^{-1}_{m,m'}$. Note that $d(\boldsymbol{\theta}_m) + d(\mathbf{u}) = d(\boldsymbol{\theta}'_{m'}) + d(\mathbf{u}')$. Green (1995) showed that if the new move is accepted as the next realization of the Markov chain with probability $a = \min\{1, r\}$, where

$$r = \frac{\pi(\mathbf{y}|m', \boldsymbol{\theta}'_{m'})\pi(\boldsymbol{\theta}'_{m'}|m')\pi(m')j(m', m)q(\mathbf{u}'|\boldsymbol{\theta}'_{m'}, m', m)}{\pi(\mathbf{y}|m, \boldsymbol{\theta}_m)\pi(\boldsymbol{\theta}_m|m)\pi(m)j(m, m')q(\mathbf{u}|\boldsymbol{\theta}_m, m, m')} |J| \tag{7}$$

with $J = \partial(\boldsymbol{\theta}'_{m'}, \mathbf{u}')/\partial(\boldsymbol{\theta}_m, \mathbf{u})$ denoting the Jacobian of the transformation, then the chain satisfies the condition of detailed balance and has the required limiting distribution $\pi(m, \boldsymbol{\theta}_m|\mathbf{y})$. The condition of detailed balance requires that the equilibrium probability of moving from a state $(m, \boldsymbol{\theta}_m)$ to $(m', \boldsymbol{\theta}'_{m'})$ equals to that of moving from $(m', \boldsymbol{\theta}'_{m'})$ to $(m, \boldsymbol{\theta}_m)$; for details, see Green (1995).

To implement the reversible-jump MCMC we need to specify the probability $j(m, m')$ for every proposed move, the proposal distributions $q(\mathbf{u}|\boldsymbol{\theta}_m, m, m'), q(\mathbf{u}'|\boldsymbol{\theta}'_{m'}, m', m)$, and the function $g_{m,m'}$. These choices do not affect the results but may be crucial for the convergence rate of the Markov chain. For example, we have found that, in the reversible-jump MCMC algorithm for the EGARCH models, the choice of $d(\mathbf{u}') = 0$ and $\boldsymbol{\theta}'_{m'} = (\boldsymbol{\theta}_m, \mathbf{u})$ (in this case $g_{m,m'}$ is the identity function), which has been proven very efficient for the log-linear models (Dellaportas and Forster 1999), turned out to be very inefficient here. The reason is that when the proposed move refers to nested models, the parameters that are "common" to both models change dramatically, and therefore the associated parameter values of $\boldsymbol{\theta}'_{m'}$ correspond to regions of low posterior probability. The immediate consequence is that the algorithm jumps with low probability between models, thus reducing the convergence rate.

Our proposed strategy, which we have successfully implemented in a series of experiments with GARCH and EGARCH models, is as follows. First, we suggest that all the parameters of the proposed model are generated from a proposal distribution. Consequently, $(\boldsymbol{\theta}'_{m'}, \mathbf{u}') = (\mathbf{u}, \boldsymbol{\theta}_m)$ with $d(\boldsymbol{\theta}_m) = d(\mathbf{u}'), d(\boldsymbol{\theta}'_{m'}) = d(\mathbf{u}), q(\mathbf{u}|\boldsymbol{\theta}_m, m, m') = q(\mathbf{u}|m'), q(\mathbf{u}'|\boldsymbol{\theta}'_{m'}, m', m) = q(\mathbf{u}'|m)$, and the Jacobian term in (7) is 1. In this case, the probability of acceptance of the new move as the next realization of the Markov chain is given by $a = \min\{1, r\}$, where

$$r = \frac{\pi(\mathbf{y}|m', \boldsymbol{\theta}'_{m'})\pi(\boldsymbol{\theta}'_{m'}|m')\pi(m')j(m', m)q(\mathbf{u}'|m)}{\pi(\mathbf{y}|m, \boldsymbol{\theta}_m)\pi(\boldsymbol{\theta}_m|m)\pi(m)j(m, m')q(\mathbf{u}|m')}. \tag{8}$$

The proposal densities $q(\mathbf{u}|m')$ and $q(\mathbf{u}'|m)$ can be chosen by investigation of a "pilot run": We start a Markov chain for each model from the best available starting values (e.g., the maximum likelihood estimates if they exist) and simulate the "within-model" Markov chain many times to obtain approximate marginal posterior means and covariance matrices for each model parameter vector. These estimates are then used to construct proposal densities $q(\mathbf{u}|m')$ and $q(\mathbf{u}'|m)$ taken as multivariate normal densities. To complete specification of our reversible-jump MCMC algorithm, we need to specify the probabilities $j(m, m')$. We have used $j(m, m') = (|M| - 1)^{-1}$ for all $m, m' \in M$, which is the simplest choice, where $|M|$ is the number of different models that are used in the reversible-jump MCMC algorithm. A sensible alternative to define $j(m, m')$ is to use approximations derived from a criterion such as the BIC, but we found that this does not work well because the resulting $j(m, m')$ values are very close to 1 for one of the models.

As an example, suppose we deal with only two models—a GARCH(1,1) model, denoted by m_1, with parameter vector $\boldsymbol{\theta}_1 = (\alpha_0, \beta_1, \sigma_0^2, \alpha_1)$, and an EGARCH(1,1) model, denoted by m_2, with parameter vector $\boldsymbol{\theta}_2 = (\tilde{\alpha}_0, \tilde{\beta}_1, \tilde{\sigma}_0^2, \tilde{\theta}_1, \tilde{\gamma}_1, \tilde{v})$. We shall illustrate in turn two reversible-jump strategies. For convenience, assume that $j(m_1, m_2) = j(m_2, m_1) = 1$; that is, we always propose to move from one model to the other. For the first strategy, we need a three-dimensional invertible function g that transforms the "common" elements of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$:

- To move from GARCH(1,1) to EGARCH(1,1),

1. set $\mathbf{u} = (\tilde{\theta}_1, \tilde{\gamma}_1, \tilde{v})$ and $\mathbf{u}' = \alpha_1$ [so $d(\mathbf{u}) + d(\boldsymbol{\theta}_1) = d(\mathbf{u}') + d(\boldsymbol{\theta}_2) = 7$],
2. choose proposal densities $q(\mathbf{u}|\boldsymbol{\theta}_1, m_1, m_2)$ and $q(\mathbf{u}'|\boldsymbol{\theta}_2, m_2, m_1)$,
3. generate $\mathbf{u}$ from $q(\mathbf{u}|\boldsymbol{\theta}_1, m_1, m_2)$,
4. set $\boldsymbol{\theta}'_2 = (g(\alpha_0, \beta_1, \sigma_0^2), \mathbf{u})$ for some invertible function g,
5. set $\boldsymbol{\theta}_2 = \boldsymbol{\theta}'_2$ and move to model 2 with probability $a = \min\{1, r\}$, where r is given by (7).

- To move from EGARCH(1,1) to GARCH(1,1),

1. set $\mathbf{u} = (\alpha_1)$ and $\mathbf{u}' = (\tilde{\theta}_1, \tilde{\gamma}_1, \tilde{v})$ [so $d(\mathbf{u}) + d(\boldsymbol{\theta}_2) = d(\mathbf{u}') + d(\boldsymbol{\theta}_1) = 7$],
2. choose proposal densities $q(\mathbf{u}|\boldsymbol{\theta}_2, m_2, m_1)$ and $q(\mathbf{u}'|\boldsymbol{\theta}_1, m_1, m_2)$,
3. generate $\mathbf{u}$ from $q(\mathbf{u}|\boldsymbol{\theta}_2, m_2, m_1)$,
4. set $\boldsymbol{\theta}'_1 = (g^{-1}(\tilde{\alpha}_0, \tilde{\beta}_1, \tilde{\sigma}_0^2), \mathbf{u})$,
5. set $\boldsymbol{\theta}_1 = \boldsymbol{\theta}'_1$ and move to model 1 with probability $a = \min\{1, r\}$, where r is given by (7).

The second strategy that we propose in such problems does not require a function g but requires proposal densities q of higher dimension:

- To move from GARCH(1,1) to EGARCH(1,1),

1. set $\mathbf{u} = (\tilde{\alpha}_0, \tilde{\beta}_1, \tilde{\sigma}_0^2, \tilde{\theta}_1, \tilde{\gamma}_1, \tilde{v})$ and $\mathbf{u}' = (\alpha_0, \beta_1, \sigma_0^2, \alpha_1)$ [so $d(\mathbf{u}) + d(\boldsymbol{\theta}_1) = d(\boldsymbol{\theta}_2) + d(\mathbf{u}') = 10$],
2. choose proposal densities $q(\mathbf{u}|m_2)$ and $q(\mathbf{u}'|m_1)$,
3. generate $\mathbf{u}$ from $q(\mathbf{u}|m_2)$,
4. set $\boldsymbol{\theta}'_2 = \mathbf{u}$,
5. set $\boldsymbol{\theta}_2 = \boldsymbol{\theta}'_2$ and move to model 2 with probability $a = \min\{1, r\}$, where r is given by (8).

- To move from EGARCH(1,1) to GARCH(1,1),

1. set $\mathbf{u} = (\alpha_0, \beta_1, \sigma_0^2, \alpha_1)$ and $\mathbf{u}' = (\tilde{\alpha}_0, \tilde{\beta}_1, \tilde{\sigma}_0^2, \tilde{\theta}_1, \tilde{\gamma}_1, \tilde{v})$ [so $d(\mathbf{u}) + d(\boldsymbol{\theta}_2) = d(\boldsymbol{\theta}_1) + d(\mathbf{u}') = 10$],
2. choose proposal densities $q(\mathbf{u}|m_1)$ and $q(\mathbf{u}'|m_2)$,
3. generate $\mathbf{u}$ from $q(\mathbf{u}|m_1)$,
4. set $\boldsymbol{\theta}'_1 = \mathbf{u}$,
5. set $\boldsymbol{\theta}_1 = \boldsymbol{\theta}'_1$ and move to model 1 with probability $a = \min\{1, r\}$, where r is given by (8).

2.3 Bayesian Model Averaging for Volatility Prediction

In time-varying volatility models such as GARCH and EGARCH, prediction of the future volatility is of particular interest. Having been able to calculate the posterior probabilities of each model, it seems natural to account for this uncertainty in our predictive inferences. Rather than choosing a single "best" model and then making inferences as if the selected model was the true model, we can use the following model-averaging approach, which provides composite predictions. Suppose that we are interested in σ^2_{T+1}, the predictive volatility at time $T+1$. Then, its posterior distribution given data $\mathbf{y}$ is given by

$$\pi(\sigma^2_{T+1}|\mathbf{y}) = \sum_{m\in M} \pi(\sigma^2_{T+1}|m,\mathbf{y})\pi(m|\mathbf{y}), \qquad (9)$$

which is an average of the posterior predictive distribution under each model weighted by their posterior model probabilities. For discussion of the preceding approach, as well as evidence that accounting for model uncertainty improves predictive performance, see Raftery, Madigan, and Hoeting (1997), Kass and Raftery (1995), Draper (1995), Madigan, Gavrin, and Raftery (1995), and Volinsky, Madigan, Raftery, and Kronmal (1997).

Computation of (9) is straightforward after the reversible-jump MCMC sampling algorithm has been implemented. First, given a model m, a posterior sample of $\pi(\sigma^2_{T+1}|m,\mathbf{y})$ is just obtained by calculating, for each sampled point in $\boldsymbol{\theta}$, the variances $\sigma^2_0,\sigma^2_1,\ldots,\sigma^2_{T+1}$ in (2) or (3). Then (9) suggests that, to obtain a sample of $\pi(\sigma^2_{T+1}|\mathbf{y})$, each sampled point under model m should be taken with probability $\pi(m|\mathbf{y})$. Thus, the derived sample of $\pi(\sigma^2_{T+1}|\mathbf{y})$ is obtained by weighting all samples of $\pi(\sigma^2_{T+1}|m,\mathbf{y})$ by the corresponding $\pi(m|\mathbf{y})$. We emphasize that we advocate the preceding procedure when the model-selection goal is prediction rather than the understanding of the data-generation mechanism. Moreover, all probability calculations are conditioned on a set of possible models, M, and therefore, $\pi(m|\mathbf{y})$ should be interpreted with care. In fact, we just average, using appropriate weights, over different available mechanisms that generate predictive distributions.

3. A SIMULATION STUDY

In this section we perform a simulation study to illustrate the relative merits of Bayesian versus classical approaches. We focus on one of the most important practical issues related to heteroscedastic models, the construction of volatility predictions. We first simulate data $y_1, y_2, \ldots, y_{1,025}$ from a GARCH(1,1) model of the form

Table 1. Mean Absolute Percentage Error for GARCH and EGARCH Models

Model	GARCH	EGARCH
Classical	.0332	.0468
Posterior mean	.0318	.0593
Posterior median	.0308	.0581

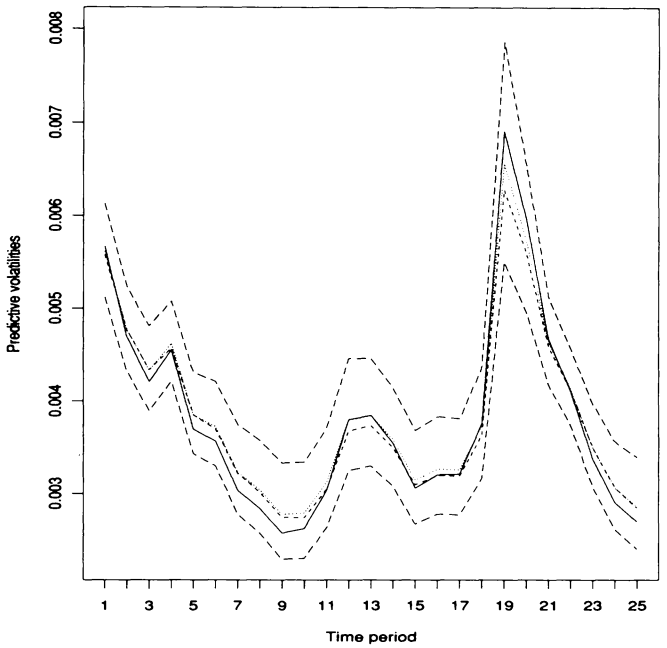

Figure 1. GARCH(1,1) Model: ———, True Volatilities; - - - -, Posterior Mean; – – –, Classical Volatilities; — — —, .025–.975 Quantile.

$$\sigma^2_t = .0005 + .25y^2_{t-1} + .7\sigma^2_{t-1}, \qquad \sigma^2_0 = .0003,$$

$$y_t \sim \mathrm{N}(0,\sigma^2_t), \qquad (10)$$

and from an EGARCH(1,1) model of the form

$$\ln(\sigma^2_t) = -.3 + .7\ln(\sigma^2_{t-1}) + .1\,\frac{y_{t-1}}{\sigma_{t-1}} + .5\left(\left|\frac{y_{t-1}}{\sigma_{t-1}}\right| - E\left|\frac{y_{t-1}}{\sigma_{t-1}}\right|\right)$$

$$\sigma^2_0 = .003, \quad y_t \sim \mathrm{GED}_2(0,\sigma^2_t),$$

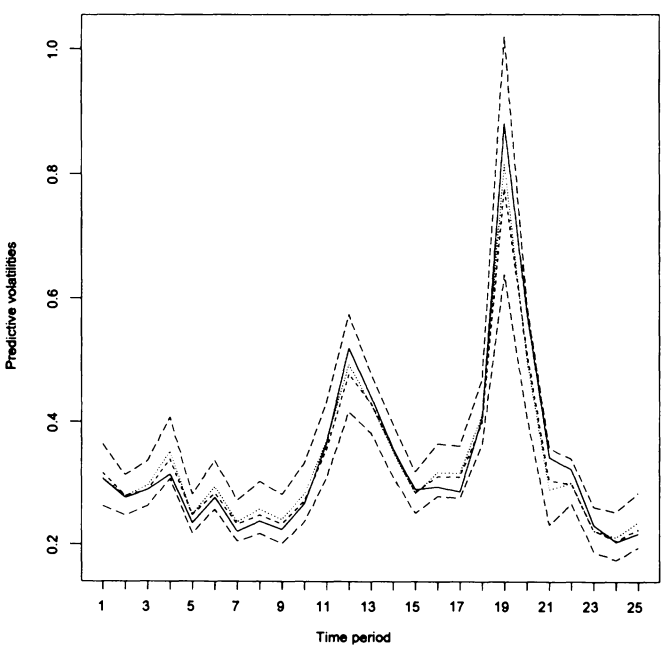

Figure 2. EGARCH(1,1) Model: ———, True Volatilities; - - - -, Posterior Mean; – – –, Classical Volatilities; — — —, .025–.975 Quantile.

Table 2. Geweke's Convergence z Scores for the Parameters of the GARCH and EGARCH Models

Model	α_0	α_1	β_1	σ_0^2	n				
GARCH(1, 1): Normal errors	1.22	.146	−.471	.934	—				
GARCH(1, 1): Student-t errors	−1.56	−.238	.667	.576	−.00276				
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	σ_0^2
EGARCH(1, 1)	.226	.190	.152	−.939	−1.180	—	—	—	1.420
EGARCH(1, 2)	.063	−.328	−.126	1.020	−1.670	.353	—	—	−.114
EGARCH(2, 1)	1.070	.993	−.253	−1.040	−.607	—	−.361	.0564	−.473
EGARCH(2, 2)	.692	−.028	.666	.102	1.180	.111	−1.000	−.409	−1.350
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	
EGARCH*(1, 1)	1.6400	1.600	−1.140	−1.330	.214	—	—	—	
EGARCH*(1, 2)	−.0318	−.976	1.070	−1.590	−.806	.899	—	—	
EGARCH*(2, 1)	1.4600	1.420	−1.600	.201	−1.750	—	.170	−1.330	
EGARCH*(2, 2)	.0921	−.651	.255	−.648	−.711	.749	−.120	1.470	

where GED_2 denotes the generalized error distribution with mean 0, variance σ_t^2, and tail-thickness parameter $v = 2$. Our simulated hypothetical scenario consists of comparisons on a daily basis of the last 25 "true" volatilities with the classical and Bayesian volatility predictions. Thus, the predictions $\hat{\sigma}_{t+1}^2$ for $t = 1{,}000, \ldots, 1{,}024$ are based on the previous t points.

A first simple criterion to compare the two prediction approaches is to use some point estimates $\hat{\sigma}_{t+1}^2$ and construct the mean absolute percentage error

$$E = \frac{1}{25} \sum_{t=1,001}^{1,025} \frac{|\sigma_t^2 - \hat{\sigma}_t^2|}{\sigma_t^2},$$

where σ_t^2 is the "true" volatility obtained from the simulated data. Table 1 provides the values of E for both GARCH and EGARCH models and for both posterior mean and medians taken as Bayesian point estimators. The classical estimators are based on the function $garch(\ )$ of S-PLUS. It is evident that the differences between the two predictive approaches are, in general, minimal.

A better insight into the performance of the estimators can be achieved by a graphical inspection of the true and the predicted volatilities. Moreover, this exercise reveals the comparative merits of the classical and Bayesian approaches. On the one hand, construction of 95% posterior credible intervals is straightforward by just calculating the .025 and .975 quantiles of the predictive density sample. On the other hand, classical estimates of the dispersion of $\hat{\sigma}_{t+1}^2$ seem to be unavailable in the literature. In any case, it is reassuring that the 95% Bayesian credible intervals for the predictive volatility contain the "true" volatilities for both GARCH and EGARCH models and for all time periods; see Figures 1 and 2.

The preceding promising aspect of Bayesian inference provides a further advantage for real financial applications; for example, the value of a call option C on a nondividend-

Table 3. Estimated Posterior Means and Standard Deviations for the Parameters of the GARCH and EGARCH Models

Model	$\alpha_0 \cdot 10^{-3}$	α_1	β_1	$\sigma_0^2 \cdot 10^{-2}$	n				
$\text{GARCH}_N(1, 1)$	.13	.31	.66	.32	—				
	.06	.08	.08	.51	—				
$\text{GARCH}_T(1, 1)$	.10	.26	.71	.45	6.59				
	.06	.07	.08	.63	2.14				
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	σ_0^2
EGARCH(1, 1)	−.46	.93	.08	.41	1.40	—	—	—	.01
	.24	.04	.04	.10	.12	—	—	—	.04
EGARCH(1, 2)	−.54	.55	.12	.52	1.43	.37	—	—	.02
	.23	.19	.05	.10	.13	.18	—	—	.07
EGARCH(2, 1)	−.29	.96	.27	.51	1.49	—	−.21	−.16	.01
	.14	.02	.08	.11	.14	—	.08	.10	.01
EGARCH(2, 2)	−.33	.89	.26	.51	1.48	.06	−.20	−.12	.01
	.19	.21	.08	.11	.14	.19	.14	.14	.03
Model	α_0	β_1	θ_1	γ_1	v	β_2	θ_2	γ_2	
EGARCH*(1, 1)	−1.03	.84	.13	.57	1.37	—	—	—	
	.37	.05	.06	.11	.12	—	—	—	
EGARCH*(1, 2)	−.92	.46	.15	.61	1.42	.40	—	—	
	.28	.15	.06	.09	.12	.16	—	—	
EGARCH*(2, 1)	−.73	.89	.28	.58	1.44	—	−.20	−.08	
	.30	.04	.08	.11	.13	—	.08	.11	
EGARCH*(2, 2)	−.99	.43	.21	.57	1.43	.42	−.08	.14	
	.40	.26	.10	.11	.13	.23	.12	.16	

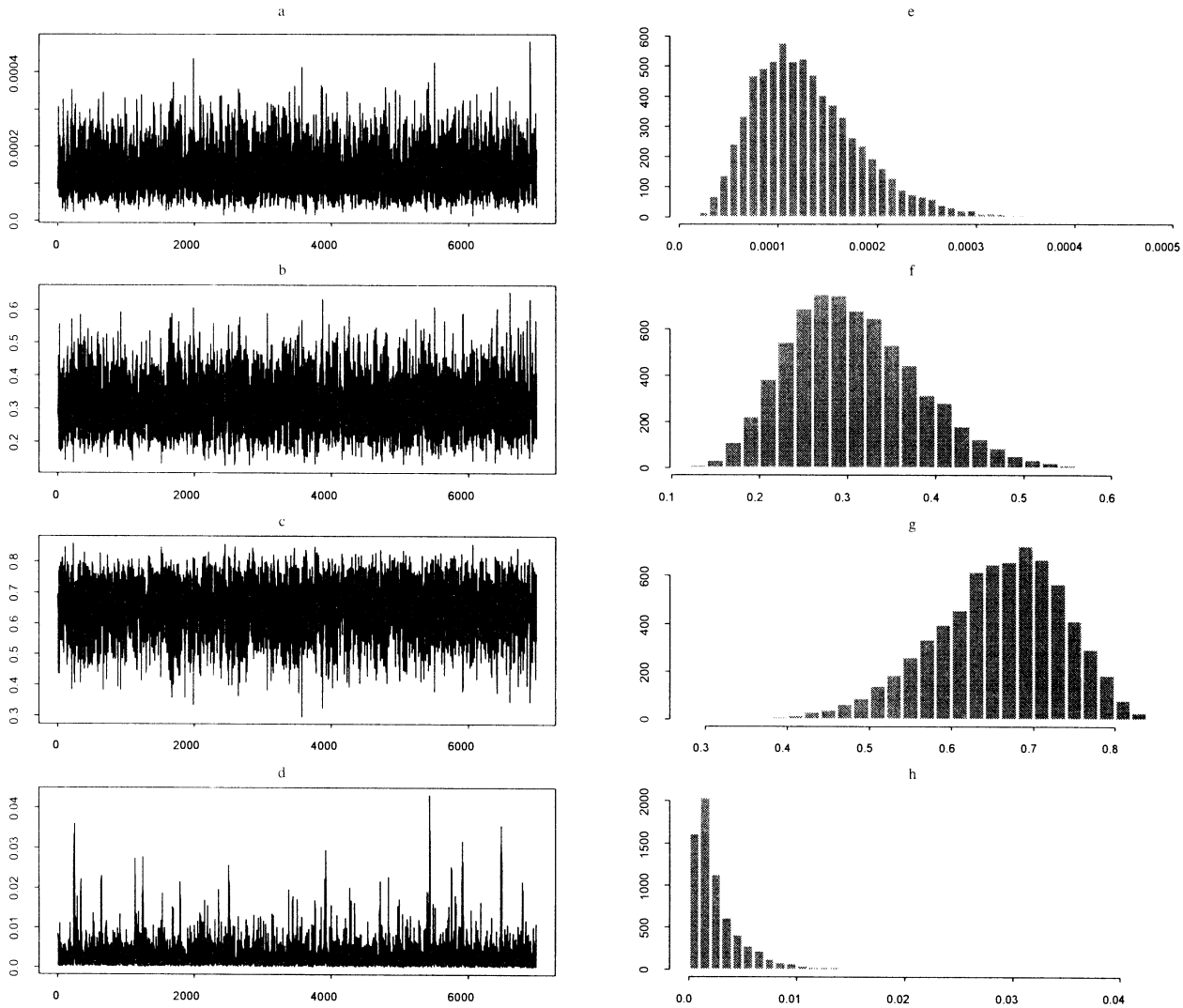

Figure 3. Convergence Diagrams and Histograms of the Posterior Sample of the Parameters of the GARCH(1,1) Model: a–d, Convergence Diagrams for α_0, α_1, β_1, σ_0^2; e–h, Histograms of the Posterior Sample of α_0, α_1, β_1, σ_0^2.

paying stock is given (e.g., see Dubofsky 1992) by $C = S\Phi(d_1) - ke^{-rt}\Phi(d_2)$, where $\Phi(.)$ is the cumulative standard normal distribution function, S is the price of the underlying asset, k is the strike price of the call option, $d_1 = (\ln(S/k) + (r + \sigma^2/2)/T)/\sigma\sqrt{T}, d_2 = d_1 - \sigma\sqrt{T}, r$ is the risk-free rate, T is the time to expiration, and σ is the standard deviation of the underlying asset's return at time to expiration. The volatility σ^2 of the stock is an important determinant of an option's value and is the only determinant of the value of the call option that is not directly observable. The prediction obtained from an ARCH-type model may be used as an estimate for σ^2 (Noh, Engle, and Kane 1994). It is immediately evident that we can utilize the predictive density of σ^2, using GARCH models and construct the posterior density of C. Moreover, if the shape of $\pi(\sigma^2|\mathbf{y})$ is not normal-like, then this feature will be realistically represented in the distribution of C.

4. AN APPLICATION: THE ATHENS STOCK–EXCHANGE INDEX

We illustrate our proposed methodology using $T = 490$ weekly rates of the General Index of the Athens stock exchange over the period 1986–1996. If G_t is the value of the General Index at time t, then we model the weekly rate $y_t = \ln(G_t/G_{t-1}), t = 1, \ldots, T$.

First, we apply GARCH(1,1) and EGARCH$(p, q), p, q = 1, 2$, models to the Athens stock-exchange data. For our illustration we chose noninformative priors for all model parameters. For the parameters of the GARCH model, we used $\pi(\alpha_0) = \alpha_0^{-1}, \pi(\sigma_0^2) = \sigma_0^{-2}$, and $U(0, 1)$ priors for α_1 and β_1. Under the Student-t distribution, the degree of freedom n $(n > 2)$ is a parameter to be estimated, and we used as a prior the noninformative $\pi(n) = (n - 2)^{-1}$. Stationarity conditions impose that $\alpha_1 + \beta_1 < 1$ for the GARCH(1,1) model; this was taken into account by just rejecting, in the MCMC algorithm, all pairs of (α_1, β_1) that did not obey the preceding restriction; for example, see Gelfand, Smith, and Lee (1992). For the parameters of the EGARCH models, we used $U(-1, 1)$ priors for $\beta_j, j = 1, 2$, and normal or lognormal noninformative priors for the other parameters of the model taken as N(0, 10) for $\alpha_0, \theta_i, \gamma_i, i = 1, 2, \text{LN}(1.04 \cdot 10^{22}, 2.93 \cdot 10^{87})$ for v, and $\text{LN}(4.72 \cdot 10^{18}, 6.008 \cdot 10^{80})$ for σ_0^2, where $\text{LN}(\mu, \sigma^2)$ denotes the lognormal distribution with mean μ and variance

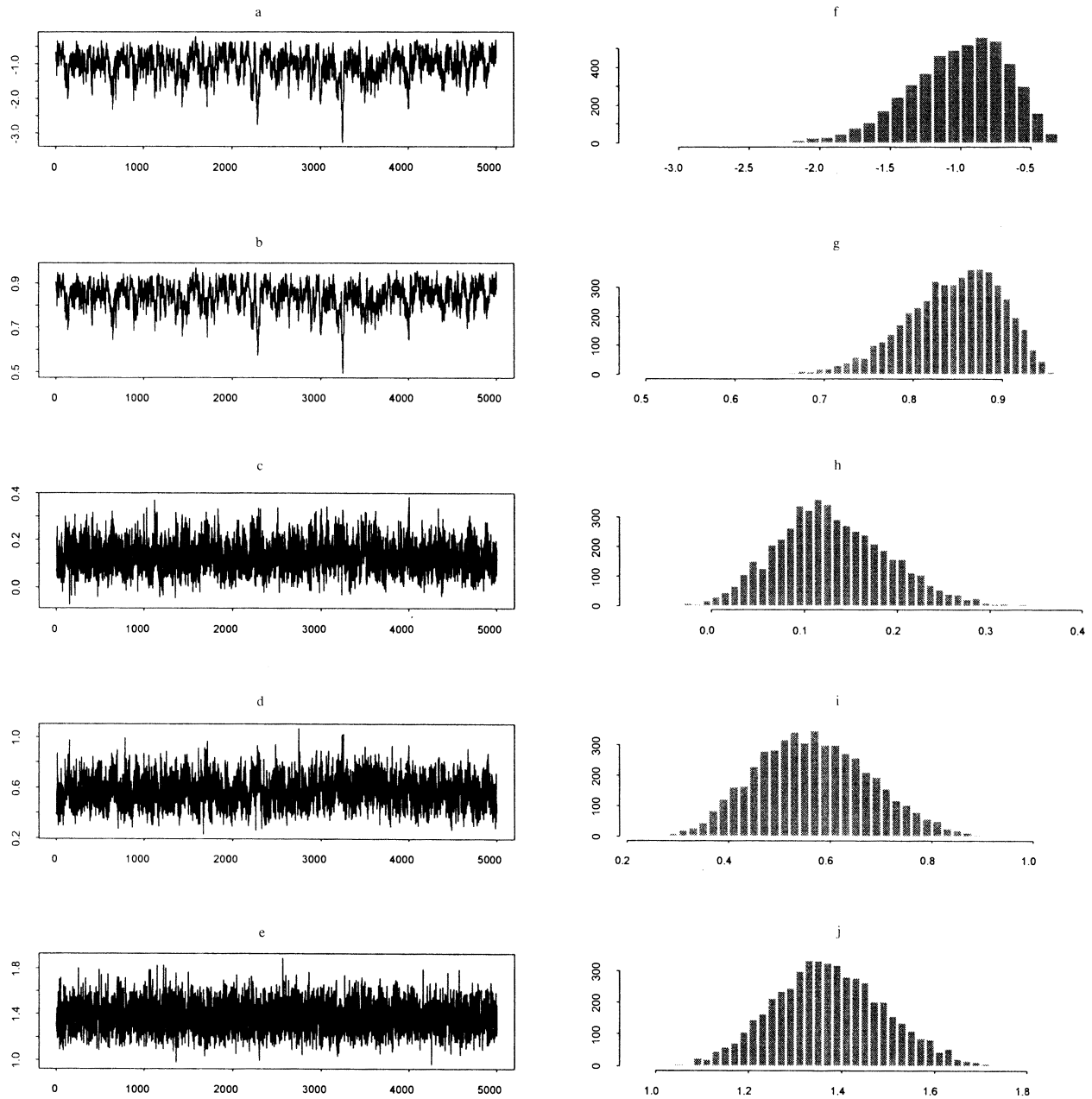

Figure 4. Convergence Diagrams and Histograms of the Posterior Sample of the Parameters of the EGARCH (1,1) Model. a–e, Convergence Diagrams for α_0, β_1, θ_1, γ_1, ν; f–j, Histograms of the Posterior Sample of α_0, β_1, θ_1, γ_1, ν.*

σ^2. The priors for both models turned out to be practically noninformative because their effective range is from 2.5 to 70 times larger than the effective range of the resulting posterior densities. Stationarity conditions also were taken into account as in the GARCH model. We use two EGARCH specifications: The first, denoted by EGARCH(p, q), is the model formulation given by (4). The second, denoted by EGARCH$^*(p, q)$, is an EGARCH model with σ_0^2 estimated from the unconditional variance of the sample data.

The output sample of every MCMC run was constructed as follows. First a large sample was taken and an initial (burn-in) part of it was discarded after a visual inspection of the time series plots of each parameter. Then the autocorrelation function of each parameter was investigated, and a decision was made about the lag intervals with which the sample should be collected to achieve a nearly noncorrelated sample. And finally, the resulting samples were checked for convergence by using the tests proposed by Geweke (1992) and Heidelberger and Welch (1983). Table 2 presents the z scores from the former diagnostic. These z scores indicate that the convergence of the Markov chain has been achieved. Estimated posterior means and standard deviations for the parameters of each model we considered are illustrated in Table 3. We present in Figures 3 and 4 the convergence diagrams and the histograms of the posterior sample of the parameters of the GARCH(1,1) and EGARCH*(1,1) models, respectively. The shapes of the posterior distribution of all the GARCH(1,1) parameters (Fig. 3) and of the α_0 and β_1 parameters in the EGARCH*(1,1) model [Fig. 4(f, g)] indicate deviation from normality.

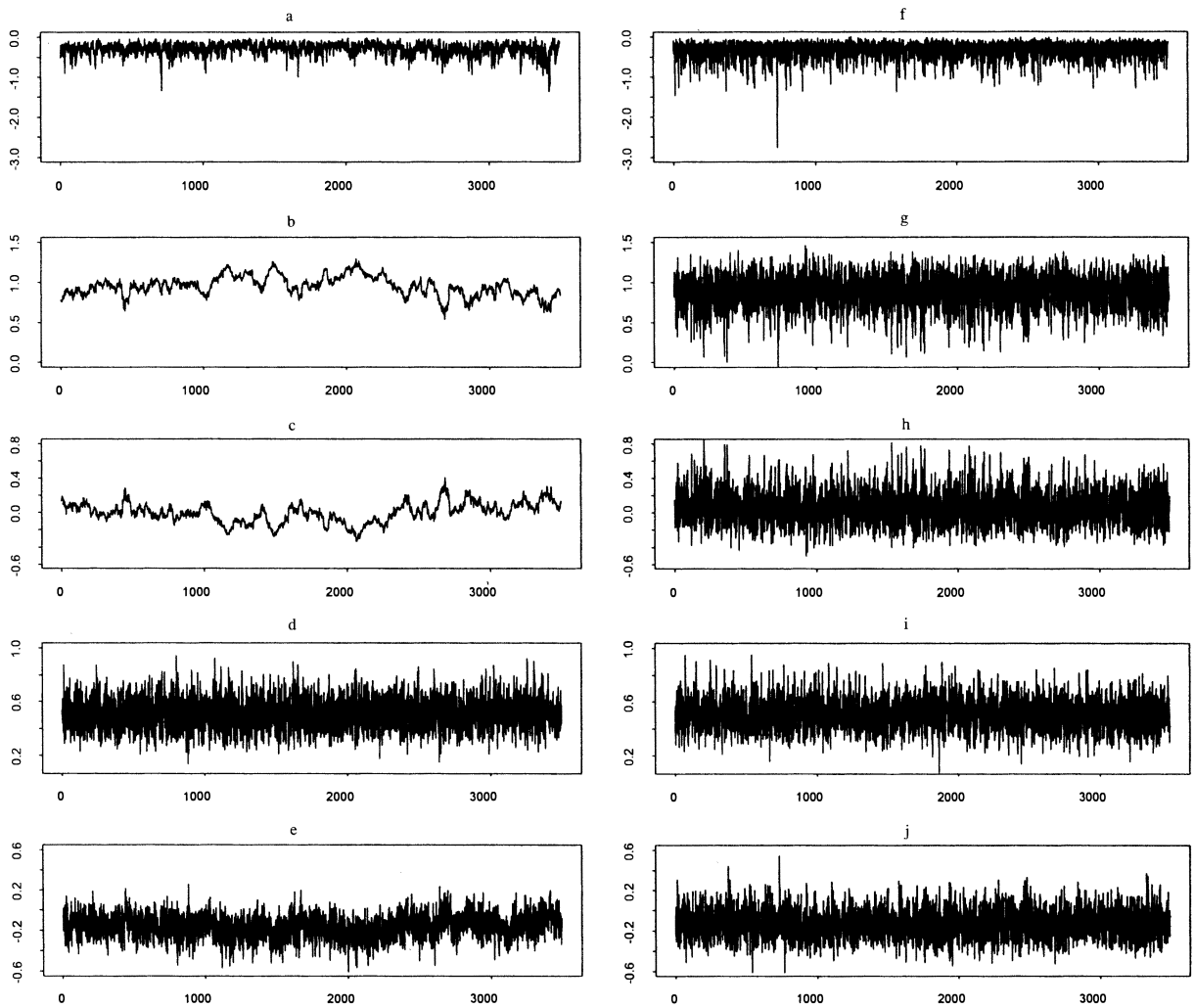

Figure 5. MCMC Output of the Highly Correlated Parameters of the EGARCH(2,2) Model. a–e, Single-Component Update, Parameters α_0, β_1, β_2, γ_1, γ_2; f–j, Simultaneous Vector Update, Parameters α_0, β_1, β_2, γ_1, γ_2.

The preceding strategy enabled us to investigate more deeply the MCMC behavior of GARCH/EGARCH models. Indeed, it soon became evident that, in some models, pairs of parameters exhibit strong posterior correlation. As is well known (Hills and Smith 1992), this phenomenon reduces the performance of the MCMC algorithm. To deal with this, we used (for the highly correlated parameters) the Metropolis multivariate step described in Section 2.1. As an example, we report that for the EGARCH(2,2) model we used the preceding strategy for the parameters $\alpha_0, \beta_1, \beta_2, \gamma_1$, and γ_2, with $c = .90$, and we achieved a reduction in the required number of iterations of at least 30%. In Figure 5, we illustrate the MCMC output of the parameters $\alpha_0, \beta_1, \beta_2, \gamma_1$, and γ_2 of the EGARCH(2,2) model using a single-component update [Fig. 5(a)–(e)] and a simultaneous vector update [Fig. 5(f)–(j)]. The results are based on 3,500 iterations taken with lag 200 and illustrate the improvement achieved with the multivariate Metropolis algorithm.

Table 4. Posterior Probabilities and Bayes Factors of Eight Competing EGARCH Models

Model	Posterior probability	Bayes factor
EGARCH(2, 1)	.4772	210.5118
EGARCH(1, 2)	.1786	78.7814
EGARCH(1, 1)	.1579	69.6343
EGARCH(2, 2)	.1066	47.0078
EGARCH*(1, 2)	.0620	27.3466
EGARCH*(1, 1)	.0128	5.6520
EGARCH*(2, 2)	.0026	1.1508
EGARCH*(2, 1)	.0023	1.0000

Our model-selection exercise consists of specifying the order of an EGARCH model as well as testing whether the (usual) prespecification of σ_0^2 affects the inferences made. For the eight competing models, we applied the reversible-jump MCMC algorithm, and the posterior probabilities in each model are illustrated in Table 4, together with the Bayes factors of all models against the (least probable) model EGARCH*(2,1). Based on these results, the "best" model for the Athens stock market is the EGARCH(2,1) model with posterior probability .4772. Proposal densities $q(\mathbf{u}|m'), q(\mathbf{u}'|m)$ for each parameter were constructed by using the MCMC output of separate model runs described earlier. These proposals are taken as multivariate normal densities with mean vector consisting of the sample mean values and covariance matrix equal to the corresponding sample covariance matrix of the parameters in each model. The behavior of the MCMC chain was good with rapid con-

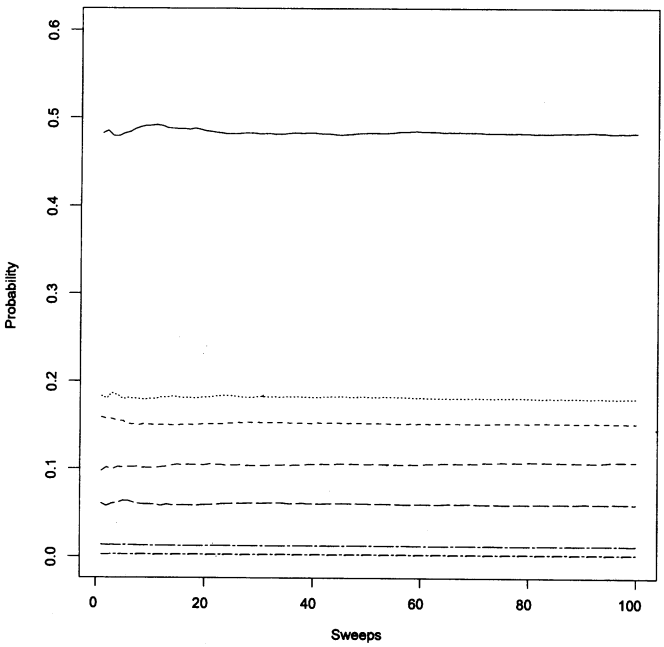

Fig. 6. Convergence Behavior of the EGARCH Models: ———, EGARCH(2, 1); - - - -, EGARCH(1,2); – – –, EGARCH(1,1); ——–, EGARCH(2,2); ——–, EGARCH(1,2); —·—·, EGARCH*(1,1); ——·, EGARCH*(2,2); - - -, EGARCH*(2,1).*

vergence of the probabilities $\pi(m|\mathbf{y})$. We ran the reversible-jump MCMC algorithm for 1,000,000 iterations, and in Figure 6, we illustrate the probabilities of the eight different EGARCH models across the sweeps calculated ergodically every 10,000 iterations. Due to the large number of models, the time for 1,000,000 iterations was around 8,000 minutes in a Pentium II 300 MHz with 128 MB RAM. Note, however, that this is an extremely conservative run; Figure 6 illustrates that a tenth or a fifth of the runs would provide essentially the same results. In the reversible-jump MCMC algorithm, we used the noninformative priors we suggested earlier in Section 4 for the estimation of the parameters of the EGARCH models.

To get some assessment of the robustness of the empirical results, we tried five different Cauchy(μ, σ^2) prior specifications for the model selection of the four EGARCH*(p, q) models. The Cauchy(μ, σ^2) prior was preferred by Jeffreys (1961), and the normal analog of the Cauchy(μ, σ^2) density is an N$(\mu, \pi\sigma^2/2)$ density according to Berger and Delampady (1987). Therefore, the analog of the priors we used in the previous analysis is $U(-1, 1)$ for β_1, β_2, Cauchy(0, 6.4) for $\alpha_0, \theta_i, \gamma_i, i = 1, 2$, and Cauchy(.7, 6.4) for $v' = \ln(v)$. We used Cauchy$(\mu, 30)$, Cauchy$(\mu, 50)$, Cauchy$(\mu, 100)$, Cauchy$(\mu, 500)$, and Cauchy$(\mu, 5{,}000)$ priors. We ran the reversible-jump MCMC algorithm for 1,000,000 iterations, and in Table 5 we illustrate the posterior model probabilities using these alternative prior specifications. The time needed was 2,310 minutes on a Pentium II 300 MHz with 128 MB RAM. Based on the results of Table 5, we gather that the results are robust to alternative prior specifications, and in all cases the most probable model is EGARCH*(1,2). The reversible-jump MCMC algorithm supports the simplest model EGARCH*(1,1) as the priors become more flat. This characteristic is well known in the Bayesian model-selection literature; see Lindley (1957), Bartlett (1957), Kass and Raftery (1995), and Berger and Delampady (1987), among others.

Some models are not visited very often in the reversible-jump MCMC algorithm; for example, just the four top models are visited with probability more than .92. Had we been required to obtain parameter estimates such as those in Table 3 together with posterior model probabilities, we should have tuned the reversible-jump MCMC accordingly. For example, it is straightforward to tune the Markov chain so that only a subset of models is visited by just changing the probabilities $j(m, m')$.

Having obtained estimates of the posterior probabilities of each model, we can apply a Bayesian model-averaging procedure to derive estimates of the composite posterior predictive volatility σ^2_{T+1}. In our illustration we calculated the predictive density σ^2_{T+1} based on the eight EGARCH models. To achieve this, we constructed all predictive densities $\pi(\sigma^2_{T+1}|m, \mathbf{y})$ under each model, for each sampled point in $\boldsymbol{\theta}$, and then we weighted all samples of $\pi(\sigma^2_{T+1}|m, \mathbf{y})$ by the corresponding posterior model probabilities taken from the reversible-jump MCMC algorithm. This curve, constructed by using the S-PLUS kernel-density estimation command *density*( ), is depicted in Figure 7 together with the corresponding estimates obtained by the three most probable models.

Finally, note that the need to impose stationarity conditions in a Bayesian context is not well understood and not broadly accepted. In our data application, we relaxed these conditions for our most probable model EGARCH(2,1), and we constructed the posterior predicted volatility σ^2_{T+1}. The posterior sample of the parameters gives probability (.0033) to nonstationarity (i.e., .0033 of the sampled points of β_1 are equal or greater than 1). The density of predicted volatility σ^2_{T+1} is also illustrated in Figure 7 and is almost identical with the density of the EGARCH(2,1) model when we use stationarity conditions.

5. DISCUSSION

In this article two important issues are considered. The first is related to the Bayesian inference of time-varying volatility models, which, we believe, offers some advant-

Table 5. Posterior Probabilities of the Four EGARCH Models Using Alternative Prior Specifications*

Model	Cauchy(μ, 6.4)	Cauchy(μ, 30)	Cauchy(μ, 50)	Cauchy(μ, 100)	Cauchy(μ, 500)	Cauchy(μ, 5,000)
EGARCH*(1, 1)	.16080	.16304	.16393	.17238	.18670	.22080
EGARCH*(1, 2)	.77801	.71327	.68659	.67388	.61706	.60558
EGARCH*(2, 1)	.02845	.05381	.05634	.05751	.04471	.04057
EGARCH*(2, 2)	.03274	.06988	.09315	.09624	.15154	.13306

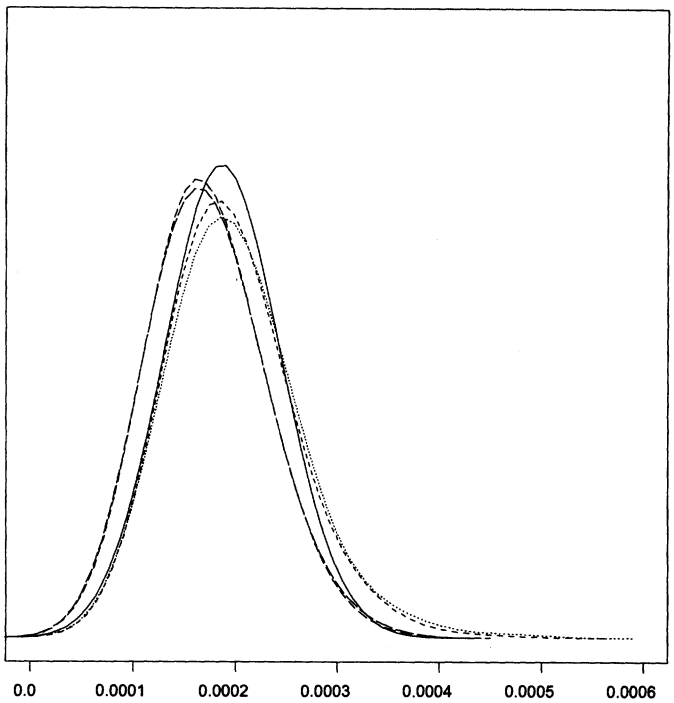

Figure 7. Posterior Density of Predictive Volatility, Using Bayesian Model Averaging (BMA), EGARCH(1,1) Model with Stationarity Conditions, EGARCH(1,2) Model with Stationarity Conditions, EGARCH(2,1) Model with Stationarity Conditions, EGARCH(2,1)⁺ Model Without Stationarity Conditions: ——, BMA; ----, EGARCH(1,1); – – – –, EGARCH(1,2); — —, EGARCH(2,1); — — –, EGARCH(2,1)⁺.

ages compared to the classical approaches. For example, local maxima do not present a problem, and posterior densities of functions of the parameters are easily available. The second is related to the model-selection problem, which is usually done via BIC or AIC. These approaches provide no direct information on the reliability of the estimates, and they do not allow prior input for model choice; moreover, their definitions and/or calibrations rely on asymptotic considerations. Our proposed strategy is to use the reversible-jump MCMC algorithm to calculate the posterior probability of every proposed model; subsequently, we allow for richer inferences through model averaging.

Finally, we note that, although we only analyzed the methodology by using GARCH and EGARCH models, our analysis can be extended to variants of these models in a straightforward way.

ACKNOWLEDGMENTS

We are very thankful to the associate editor and two referees for their constructive comments. We also thank the General Secretariat of Research and Technology of Greece for supporting this research.

[Received March 1999. Revised July 1999.]

REFERENCES

Baillie, R., and Bollerslev, T. (1989), "The Message in Daily Exchange Rates: A Conditional Variance Tale," *Journal of Business & Economic Statistics*, 7, 297–305.

Bartlett, M. S. (1957), Comment on "A Statistical Paradox," by D. V. Lindley, *Biometrika*, 44, 533–534.

Berger, J. O., and Delampady, M. (1987), "Testing Precise Hypotheses" (with discussion), *Statistical Science*, 2, 317–352.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69, 542–547.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), "ARCH Modeling in Finance—A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis–Hastings Algorithm," *The American Statistician*, 49, 327–335.

Dellaportas, P., and Forster, J. (1999), "Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Log-Linear Model," *Biometrika*, 86, 615–633.

Draper, D. (1995), "Assessment and Propagation of Model Uncertainty" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 57, 45–97.

Dubofsky, D. A. (1992), *Options and Financial Futures: Valuation and Uses*, Singapore: McGraw-Hill.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

——— (1995), *ARCH Selected Readings*, New York: Oxford University Press.

Engle, R. F., and Ng, V. K. (1993), "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1749–1778.

Evans, M., and Swartz, T. (1995), "Methods for Approximating Integrals in Statistics With Special Emphasis on Bayesian Integration Problems," *Statistical Science*, 3, 254–272.

Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling," *Journal of the American Statistical Association*, 873, 523–532.

Geweke, J. (1989), "Exact Predictive Densities for Linear Models With ARCH Disturbances," *Journal of Econometrics*, 40, 63–86.

——— (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 169–193.

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Heidelberger, P., and Welch, P. W. (1983), "Simulation Run Length Control in the Presence of an Initial Transient," *Operations Research*, 31,1109–1144.

Hills, S. E., and Smith, A. F. M. (1992), "Parameterization Issues in Bayesian Inference," in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Clarendon Press, pp. 227–246.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models" (with discussion), *Journal of Business & Economic Statistics*, 12, 371–417.

Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), London: Oxford University Press.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors and Model Uncertainty," *Journal of the American Statistical Association*, 90, 773–795.

Lindley, D. V. (1957), "A Statistical Paradox," *Biometrika*, 44, 187–192.

Madigan, D., Gavrin, J., and Raftery, A. E. (1995), "Eliciting Prior Information to Enhance the Predictive Performance of Bayesian Graphical Models," *Communications in Statistics, Part A—Theory and Methods*, 24, 2271–2292.

Muller, P., and Pole, A. (in press), "Monte Carlo Posterior Integration in GARCH Models," *Sankhyā*, 62.

Nelson, D. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.

Nobile, A., and Green, P. J. (in press), "Bayesian Analysis of Factorial Experiments by Mixture Modelling," *Biometrika*, 87.

Noh, J., Engle, R. F., and Kane, A. (1994), "Forecasting Volatility and Option Prices of the S&P 500 Index," *Journal of Derivatives*, 17–30.

Pagan, A. R., and Schwert, G. W. (1990), "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics*, 45, 267–290.

Raftery, A. E., Madigan, D., and Hoeting, J. (1997), "Model Selection and Accounting for Model Uncertainty in Linear Regression Models," *Journal of the American Statistical Association*, 92, 179–191.

Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 59, 731–792.

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, London: Chapman & Hall, pp. 1–55.

Smith, A. F. M., and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Ser. B, 55, 3–24.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.

Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997), "Bayesian Model Averaging in Proportional Hazard Models: Predicting the Risk of a Stroke," *Applied Statistics*, 46, 443–448.