Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series B, Pt. 1, pp. 127-144

MONTE CARLO POSTERIOR INTEGRATION IN GARCH MODELS*

By PETER MÜLLER
Duke University, Durham
and
ANDY POLE
Invictus Partners, New York

SUMMARY. Recent developments in estimating non-linear, non-normal dynamic models have used Gibbs sampling schemes over the space of all parameters involved in the model (Carlin, Polson and Stoffer, 1992) as well as Monte Carlo integration based on propagating a Monte Carlo sample on the parameter vector θ_t through the stages of the dynamic model. Both approaches have advantages. The first because it enables a convenient Gibbs sampler implementation. The latter because it splits the problem into a series of simulation problems, one for each time step, thus having computational effort only increase linearly with the length of the time series. Also, propagating a Monte Carlo sample through the dynamic model allows for unrestricted generality in the distributional form of evolution noise and likelihood. In this paper we develop schemes along both lines to apply to the analysis of GARCH (generalized autoregressive conditional heteroskedasticity) models for daily exchange rate data.

1. Introduction

ARCH (autoregressive conditional heteroskedasticity) models have achieved a considerable following in the econometrics and finance literature since their introduction by Engle (1982). An ARCH model is a discrete stochastic process with the characteristic feature that the variance at time t is some time-varying function of the time $t-1$ information set. Below, in (1) we give a more specific formal definition. Applications and studies were stimulated by the generalisation to GARCH models in Bollerslev (1986). There are now over 300 papers in the mainstream statistics and econometrics journals discussing theoretical properties

AMS (1991) *subject classification.* 60G35, 62F15,62P05, 90A20.
Key words and phrases. Exchange rate data; General dynamic model; Markov Chain Monte Carlo; Metropolis; State space model.
*This research was partially supported by NSF grant DMS-9305699 at Duke University.

of many formulations of GARCH model as well as numerous applications. An excellent survey of the literature is Bollerslev *et al.* (1992); some of the review material later in this section is based on that article.

The contribution of this paper is twofold. Firstly, novel combinations of Markov Chain Monte Carlo techniques are developed. Though prompted by, and developed here specifically for, dynamic GARCH models, the algorithms are useful for a wide range of non-standard sequential analyses, dynamic or not. The methods build on and are related to Markov Chain Monte Carlo techniques discussed in Carlin, Polson and Stoffer (1992), Jacquier, Polson and Rossi (1994 and 1995), and Tierney (1994). Secondly, the class of (multivariate) GARCH models is extended to a dynamic setting thereby allowing much greater modelling flexibility. We note that dynamic modelling has proved immensely useful with traditional linear regression type models, particularly in forecasting applications. From that we suggest that dynamic formulations of the GARCH model family have similarly good potential for advancing GARCH based studies. The application to exchange rates discussed in this paper lends support to this claim, although more work is required to explore covariance structures. The implementation of the dynamic model is based on methods defined in Müller (1992).

In the remainder of this introduction we review the salient applied literature to give an overview of the setting for our work and to motivate our objectives for the paper. In Section 2 we introduce the data set that will be used for illustration. In Section 3 we discuss the estimation of a simple GARCH(1,1) model by an independence Chain type Markov Chain Monte Carlo algorithm. And in Section 4 we extend the basic GARCH(1,1) model to a multivariate dynamic model which we estimate by a Monte Carlo algorithm based on simulating the dynamic model.

1.1 *Review.* The basic ARCH model introduced in Engle (1982) allows for the error terms in a time series model to have a time varying variance. Specifically the variance is directly related to previous values of the observation series such that higher series values lead to higher stochastic variance in succeeding periods. Let $\{\epsilon_t\}$ be a stochastic process with mean 0 and variance h_t,

$$\epsilon_t \overset{ind}{\sim} [0, h_t]. \qquad \ldots(1)$$

The variance is related to previous squared values of the process in an autoregressive scheme,

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \ldots + \alpha_p \epsilon_{t-p}^2.$$

This is the basic ARCH(p) model. Bollerslev (1986) extended the ARCH formulation to include lagged values of the variance itself in the variance equation thereby providing a more flexible model form. The GARCH(p,q) model for

variance is

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-j}.$$

One of the motivations for the GARCH formulation was the large numbers of lags required in many applications of ARCH. GARCH models share with classical time series ARMA models the property that some high order autoregressive models are well approximated by low order moving average models. In most applications of GARCH a first order model, GARCH(1,1), is found to be appropriate. We introduce a Markov Chain Monte Carlo scheme to estimate GARCH(1,1) models in Section 3.

Most of the literature adopts the gaussian distribution for the conditional sampling distribution,

$$\epsilon_t \sim \mathrm{N}[0, h_t].$$

However, several alternatives have also been investigated. Bollerslev (1987) suggests the student t distribution, Jorion (1988) uses a normal-Poisson mixture, Baillie and Bollerslev (1989) use the power exponential model, Hsieh (1989) uses a normal-lognormal mixture distribution, and Nelson (1990) uses the generalised exponential distribution.

The basic multivariate GARCH(p,q) model was first studied by Bollerslev, Engle, and Wooldridge (1988). Let $vech(A)$ denote the operator that stacks the lower portion of an $n \times n$ matrix A as an $n(n+1)/2 \times 1$ vector. The vector GARCH(p,q) model is formulated as

$$\begin{array}{lcl} \tilde{\epsilon}_t & \sim & [0, \Omega_t], \\ vech(\Omega_t) & = & W + \sum_{i=1}^{p} A_i vech(\tilde{\epsilon}_{t-i}\tilde{\epsilon}'_{t-i}) + \sum_{j=1}^{q} B_i vech(\Omega_{t-j}), \end{array}$$

where $\tilde{\epsilon}_t$ is an N-vector, W is an m-vector, $m = N(N+1)/2$, A_i and B_j are $m \times m$ positive definite matrices. There is a total of $m[1+mpq]$ parameters in this model; in practice this dimensionality is greatly reduced by imposing additional structural assumptions such as diagonal A_i and B_j matrices. Bollerslev (1990) has a time varying conditional covariance matrix, Ω_t, but assumes time invariant conditional correlations.

Maximum likelihood estimation is the generally favoured scheme of inference, although semi- and non-parametric techniques have also been applied by some authors, among them Gallant and Nychka (1987), Gallant and Tauchen (1989), Engle and Gonzalez-Rivera (1991), Whistler (1988). See also the review by Bollerslev *et al.* for discussion and many more references. Bayesian inference has been used by Geweke (1989).

Bollerslev *et al.* (1992) state on page 15 that "Asymptotic theory for ARCH models is notoriously difficult." In this paper we make inference using the Bayesian paradigm. Conceptually the inference problem is straightforward:

given the data, the model form, and the distributional assumptions, the posterior distributions of model parameters yield the full inferential picture. Technically, inevitably, matters are not so straightforward. The model is nonlinear and analytical results for posterior distributions are not available. We therefore utilise numerical integration schemes to compute the posteriors.

In Section 4 we propose a model that is at once a natural extension of the multivariate GARCH model into the dynamic setting, and also a natural extension of the dynamic linear model to include GARCH variance specifications. The sequential analysis of the dynamic model is problematic because of the need to reconstruct prior and posterior distributions at each iteration. This problem was first studied by Pole and West (1990) in the context of univariate nonnormal and nonlinear dynamic models not including GARCH components. The numerical scheme proposed there used a combination of dynamic grids and Gauss Hermite integration and, while effective in some problems, is rather restricted. The approach developed here uses a much more general Monte Carlo integration scheme.

2. Exchange Rate Applications

GARCH models have been applied to foreign exchange rate data by a number of authors. Several alternative distributions as well as nonparametric estimation techniques have been tried. Hsieh (1988, 1989) applies a GARCH(1,1) model to daily rates concluding that most of the nonlinear stochastic dependencies are captured by the model (Bollerslev, *et al.*, 1992, p.38). Baillie and Bollerslev (1989) apply a GARCH(1,1) model with nonnormal error distributions to daily rates. Engle and Bollerslev (1986) use a GARCH(1,1) model for weekly dollar-Swiss franc rates (on a logarithmic scale), obtaining parameter estimates indicating an almost integrated model, $\alpha_1 + \beta_1 = 0.996$. Other applications include Lee (1988), a bivariate ARCH(12) model for deutschemark and yen-dollar weekly rates; Baillie and Bollerslev (1990), a four dimensional GARCH(1,1) model for weekly one month forward rates on european currencies; Bollerslev (1990), a multivariate GARCH(1,1) model with constant conditional correlations.

In this paper we will analyze a data set of daily exchange rates for dollar-deutsche mark, dollar-Swiss franc, and dollar-yen for the period July 7, 1988 to December 31, 1992. The observations are transformed to log differences ("returns" in the finance literature). Figure 1a plots the three dimensional time series. For the sake of clarity the figure shows the log exchange rates before differencing. Figure 1b shows the reduced data set used to estimate the static GARCH(1,1) model in Section 3.

3. An Independence Chain for the GARCH (1,1) Model

3.1 *The model.* We consider the following GARCH(1,1) model:

$$\begin{aligned} y_t &= x_t'\beta + \epsilon_t, \\ \epsilon_t &\sim N(0, h_t) \\ h_t &= z_t'\gamma, \quad h_t > 0, \end{aligned}$$

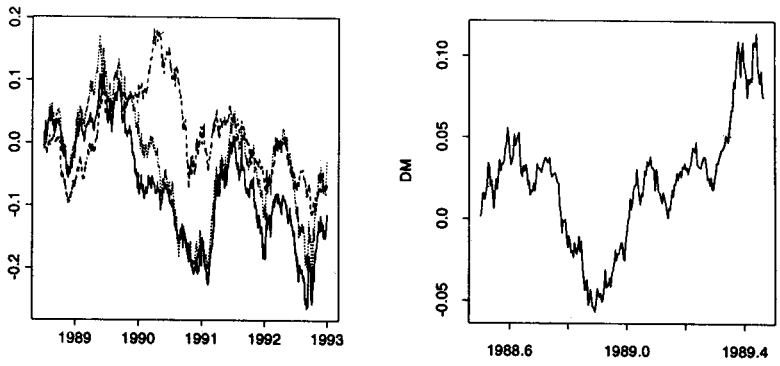


Figure 1 : Panel (a) plots the log exchange rates for German mark (solid line), Swiss franc (dotted) and yen (dashed) - all relative to 1988. Panel (b) shows the subset of the deutsche mark time series used in Section 3.

where the observations y_t are log differences of daily exchange rates for dollar-deutsche mark, recorded over the period July 7, 1988 to June 23, 1989, $x_t = (1, y_{t-1}, \mathrm{SFR}_{t-1}, \mathrm{YEN}_{t-1})$, and $z_t = (1, \epsilon_{t-1}^2, h_{t-1})$. Here SFR and YEN are the log differences of the exchange rates for dollar-Swiss franc and dollar-yen respectively. We are fitting this model to only part of the data set described earlier since the assumption of a static model would not be adequate over the extended period of the full data set. In Section 4 we discuss estimation of a dynamic model over the complete data set. Although the dynamic model offers considerably more generality, the static GARCH(1,1) model has the clear advantages of parsimony and ease of interpretation of the parameters.

3.2 *Estimation.* We estimate the model using a Markov Chain Monte Carlo algorithm following the independence chain setup of Tierney (1992). The implementation is similar to the scheme used in Jaquier, Polson and Rossi (1994). It resembles a Gibbs sampler defined by iterative draws from the conditional posteriors $p(\beta|\gamma, y)$ and $p(\gamma|\beta, y)$. However, we replace the conditional posteriors by probing distributions $g_\gamma(\gamma|\beta)$ and $g_\beta(\beta|\gamma)$ which are chosen to mimick the true conditional posteriors and to allow efficient random number generation. To still maintain the joint posterior $p(\gamma, \beta|y)$ as limiting distribution of the simulated Markov Chain an additional rejection step is required.

Let $\theta^t = (\theta_1^t, ...\theta_7^t) = (\beta_t, \gamma_t)$ denote the parameter vector after t iterations of the Markov Chain, and – for the sake of typographical ease surpressing the dependence on y – let $p(\beta, \gamma)$ denote the joint posterior. The following steps give a formal description of the implemented algorithm by defining the transition from θ^t to $\theta^{t+1}, \theta^{t+2}$ etc: (i) Define a "candidate" $\tilde{\theta} = (\beta_t, \tilde{\gamma})$ by generating $\tilde{\gamma} \sim g_\gamma(\gamma|\beta_t)$. (ii) With probability

$$a(\theta^t, \tilde{\theta}) = \min\left(1, \frac{p(\tilde{\theta})}{g_\gamma(\tilde{\gamma}|\beta_t)} \cdot \frac{g_\gamma(\gamma_t|\beta_t)}{p(\theta^t)}\right)$$

accept $\tilde{\theta}$ as new state, i.e. $\theta^{t+1} := \tilde{\theta}$, else $\theta^{t+1} := \theta^t$. (iii) Define a "candidate" $\tilde{\theta} = (\gamma_{t+1}, \tilde{\beta})$ by generating $\tilde{\beta} \sim g_\beta(\beta|\gamma_{t+1})$. (iv) With probability $a(\theta_{t+1}, \tilde{\theta})$ set $\theta^{t+2} := \tilde{\theta}$, else $\theta^{t+2} := \theta^{t+1}$. (v) Repeat Steps i through iv to simulate θ^{t+3} etc.

A potential problem in this scheme is that the acceptance probabilities a could be close to zero. This would happen when the chain encounters points where the probing distribution is significantly thinner than the true conditional posterior distribution, i.e. $g_\gamma(\gamma_t|\beta_t)/p(\theta^t)$ is close to zero. To avoid the simulated chain from getting absorbed in such points, we include a few Metropolis steps (Tierney, 1994) after every 10 iterations of the independence chain. The implemented Metropolis chain changes one coordinate at a time, scanning over all 7 coordinates. Starting with current state θ^t the chain is formally defined by the following steps: (i) Generate a "candidate" $\tilde{\theta} = (\tilde{\theta}_1, \theta_2^t ... \theta_7^t)$ by replacing the first coordinate of θ^t by $\tilde{\theta}_1 \sim N(\theta_1^t, 0.25 \cdot \sigma_1)$. (ii) With probability $a(\theta^t, \tilde{\theta}) = \min(1, p(\tilde{\theta})/p(\theta^t))$ let $\theta^{t+1} := \tilde{\theta}$, else $\theta^{t+1} := \theta^t$. (iii) Repeat Steps (i) and (ii), scanning over all coordinates. (iv) Repeat Steps (i) through (iii) four times. In Step (i), σ_i is the approximate posterior standard deviation of the i-th parameter (estimated over an initial burn-in period).

The factor $c = 0.25$ is an arbitrarily chosen scaling factor. Too large a step size (large c) would lead to almost all candidates being rejected, too small a step size, would result in a slowly mixing chain. For the given problem, the factor $c = 0.25$ turned out to be adequate in the sense that around 50% of the candidates end up being rejected. For a specific case Gelman, Roberts and Gilks (1996) show that an acceptance rate of around 25% is optimal. See Polson (1996) for a formal discussion of the underlying theoretical issues. For a practical implementation we recommend to pick a reasonable step size factor by "trial and error".

Also, the number of $n = 4$ iterations in this mini Metropolis is an arbitrary choice. Our rationale was that we did not want to change the algorithm to a Metropolis chain (hence no large n), but still wanted to leave a good chance of leaving a point with high a. With average acceptance probabilities in the mini Metropolis around 50%, a length $n = 4$ seemed reasonable.

3.3 *The probing distributions.* The choice of the probing distributions is motivated by approximating regression problems. For g_β consider the model

$y_t = x_t'\beta + \epsilon_t$ with $\epsilon_t \sim N(0, h_t)$, where the residuals that are required to evaluate $h_t = \gamma_0 + \gamma_1\epsilon_{t-1}^2 + \gamma_2 h_{t-1}$ are computed using the current values of β: Let $\hat{\beta}$ denote the current value of β. We first compute $\epsilon_{t-1} = y_{t-1} - x_{t-1}'\hat{\beta}$, and then use ϵ_{t-1} and h_{t-1} to compute $h_t = z_t'\gamma = \gamma_0 + \gamma_1\hat{\epsilon}_{t-1}^2 + \gamma_2 h_{t-1}$. The posterior distribution on β in this auxilary model provides the probing distribution $g_\beta(\beta|\gamma)$.

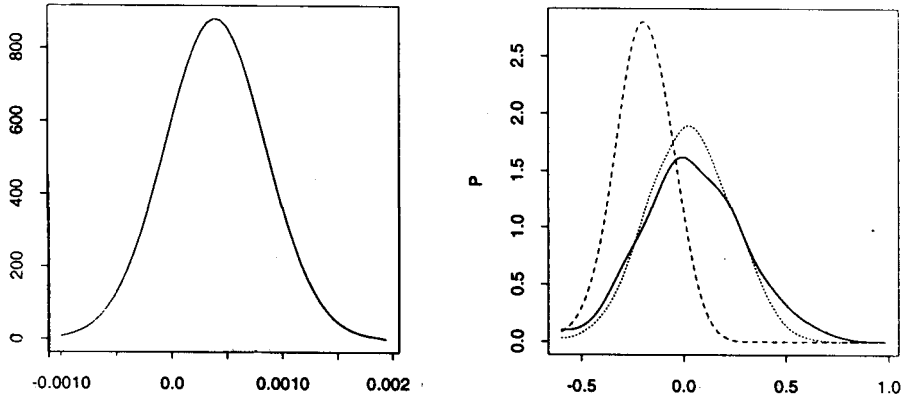
Similarily the probing distribution g_γ is derived from the following regression model which is motivated by the conditional expectation of ϵ_t^2: $\epsilon_t^2 = z_t'\gamma + \omega_t$ with $\omega_t \sim N(0, \sigma^2)$, where the γ values from the current iteration are substituted when computing the elements of z_t. Here σ is assumed unknown with a prior distribution $p(\sigma) \propto 1/\sigma$. Using the conditional variance of ϵ_t^2 in place of σ^2 did not improve the acceptance rate in Step (ii) of the independence chain.

3.4 *Results.* The following results are based on simulating 5 parallel chains implementing the earlier described Markov Chain Monte Carlo scheme. To initialize the simulation we computed for each parameter θ_i, $i = 1, \ldots, 7$, approximate posterior means $\bar{\theta}_i$ and standard deviations σ_i by Laplace integration (Tierney and Kadane, 1986) and set $\theta_i^0 \sim N(\bar{\theta}_i, 16 \cdot \sigma_i^2)$. We used multiple parallel chains with an overdispersed (relative to the approximate posterior variances σ_i^2) initial sample to be able to apply the convergence diagnostic introduced in Gelman and Rubin (1992). Gelman and Rubin propose as a convergence diagnostic the estimated factor by which the scale of the current distribution for some scalar of interest might be reduced if the simulations were continued in the limit $m \to \infty$. They introduce a statistic R which estimates this factor when $n > 1$ parallel chains are started with an overdispersed initial sample.

Table 1. ESTIMATED POTENTIAL SCALE REDUCTION FOR THE GARCH PARAMETERS

	Estimated	97.5-th Percentile
γ_1	1.0002	1.0005
γ_2	1.0000	1.0001

After 50,000 iterations we computed estimated potential scale reductions reported in Table 1. All values are very close to 1.0, indicating that posterior inference would not change noticably if we were to continue simulation. We stopped the simulation and used the Monte Carlo sample generated over the last 4000 iterations to evaluate the desired features of the posterior distribution. To reduce serial correlation we kept only every 10-th iteration, giving a total Monte Carlo sample size of $M = 2000$. In retrospect, a shorter simulation run lengths of around 10,000 iterations would have been sufficient to achieve comparable numerical accuracy. To estimate, for example, the marginal posterior distribution of parameter θ_i we used $\hat{p}(\theta_i) = M^{-1}\sum p(\theta_i|\theta_{-i}^j)$, where θ_{-i}^j denotes the j-th parameter vector from the Monte Carlo sample without the i-th coordinate. Gelfand and Smith (1990) refer to $\hat{p}$ as "Rao-Blackwellized" density estimate.



(a) $p(\beta_0|y)$ (b) $p(\beta_j|y)$

Figure 2: **Marginal posterior densities $p(\beta_0|y)$ (Panel a), $p(\beta_1|y)$ (Panel b, solid line), $p(\beta_2|y)$ (dotted line), and $p(\beta_3|y)$ (dashed line). The posterior gives high probability on a negative effect of the previous day's rate for USD/YEN.**

Figure 2 plots the estimated marginal posterior distributions for the parameters in the mean equation. While the posterior distribution for β_1 and β_2 are centered around zero, the marginal posterior $p(\beta_3|y)$ shows an interesting asymmetry. The marginal posterior distributions for the parameters γ_0, γ_1 and γ_2 are shown in Figures 3 and 4. The skewness of $p(\gamma_2|y)$ is caused by the positivity constraint on γ_0 and a high posterior correlation of γ_0 and γ_2. One-step ahead forecasts for the standard deviations $\sqrt{h_t}$ are plotted in Figure 5.

3.5 *Summary.* The suggested Markov Chain Monte Carlo algorithm allows a full Bayesian analysis of the GARCH(1,1) model without introducing the conditional variances as latent variables. If desired the independence chain could easily be adapted for a different error distribution by appropriately changing the acceptance probabilities $a(\cdot)$ only. For the simple GARCH(1,1) model we could use probing distributions based on the auxilary regression problems described in Section 3.3. Alternative implementations could be based on the iterations of MLE schemes used for GARCH models (see Bollerslev, 1986). We chose to assess convergence using the diagnostic proposed by Gelman and Rubin (1992). Alternative and complimentary schemes discussed in Geweke (1992) or Zellner and Min (1995) could be used.

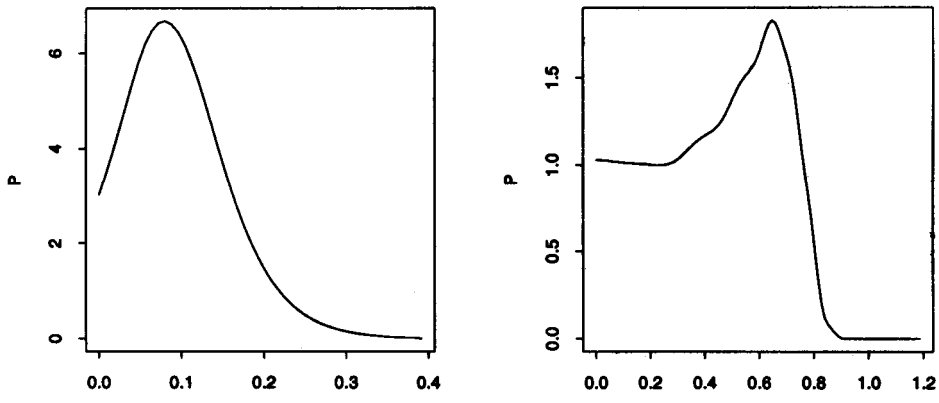


Figure 3: Marginal posterior densities $p(\gamma_1|y)$ and $p(\gamma_2|y)$. Compare with the next figure for an explaination of the skewness of $p(\gamma_2|y)$.

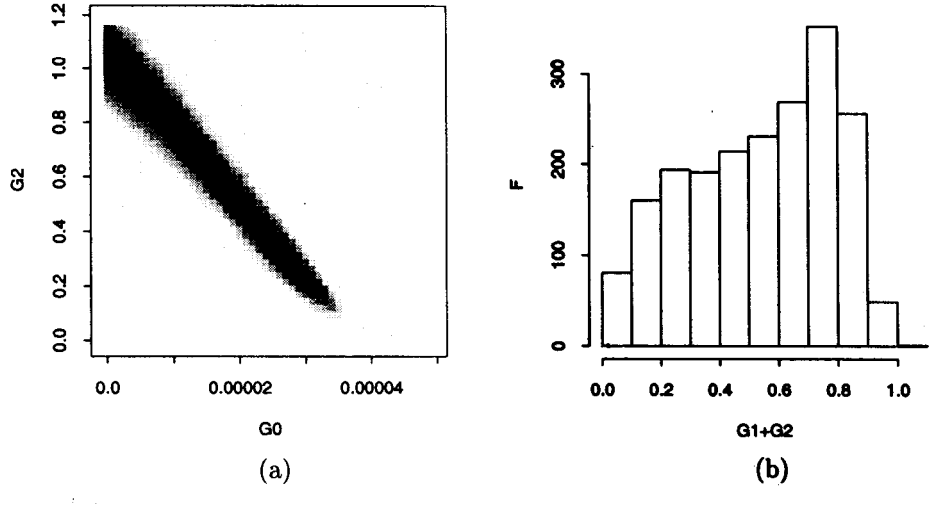


Figure 4: Posterior $p(\gamma 0, \gamma 1, \gamma 2|y)$. Panel (a) shows $p(\gamma_0, \gamma_2|y)$. The positivity constraint on γ_0 causes a posterior mode on the boundary. Together with a high posterior correlation between γ_0 and γ_2 this explains the sharp cutoff in the marginal $p(\gamma_2|y)$ shown in the previous figure. Panel (b) shows $p(\gamma_1 + \gamma_2|y)$. The posterior gives little probability to non-stationarity although $\gamma_1 + \gamma_2 < 1$ was not enforced a priori.

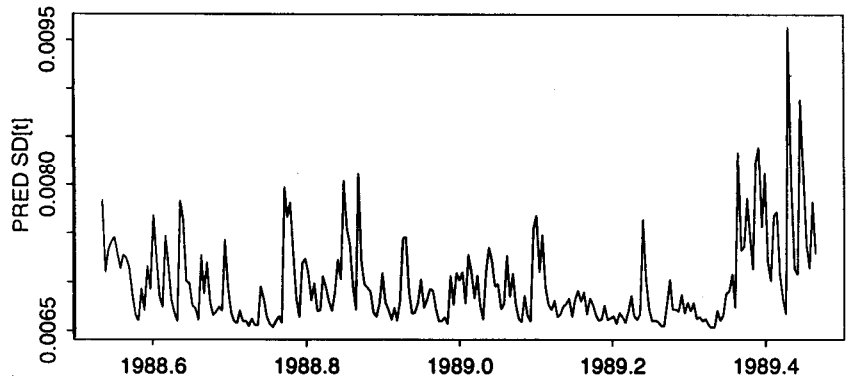


Figure 5: Predicted standard deviations σ_t ($h_t = \sigma_t^2$).

4. General Dynamic Model

4.1 *A multivariate dynamic GARCH model.* The model that we consider in this section combines elements of both the GARCH model and the multivariate dynamic linear model. The univariate GARCH model is extended by allowing covariances between the time series and introducing parameters that change smoothly over time. The usual normal dynamic linear model is extended by introducing the GARCH structure and evolution noise on the covariance matrix.

$$\begin{aligned} & \text{Observation:} \\ y_t &= a_t + A_t \cdot y_{t-1} + \epsilon_t, \\ & \epsilon_t \sim N(0, \Sigma_t) \end{aligned} \qquad \ldots(2)$$

$$\begin{aligned} & \text{Evolution:} \\ \Sigma^*_{t+1} &= W_t' W_t + B_t \Sigma_t B_t' + C_t Q_t C_t', \\ & Q_t = \textstyle\sum_{j=1}^{5} \epsilon_{t-j} \cdot \epsilon_{t-j}', \end{aligned} \qquad \ldots(3)$$

$$\begin{aligned} \theta_{t+1} &= \theta^*_t + \omega_t, \\ & \omega_t \sim N(0, \Omega_t), \end{aligned} \qquad \ldots(4)$$

where $\theta_t = vech(a_t, A_t, W_t, B_t, C_t, V_t)$ is the the vector of all parameters at time t, with V_t being the Choleski decomposition of Σ_t ($V_t' V_t = \Sigma_t$). In θ^*_t the covariance matrix Σ_t is replaced by Σ^*_{t+1}. The interpretation of the parameters in a_t and A_t is similar to the β vector in the univariate GARCH model. The matrices W_t, B_t and C_t parallel the parameters γ_0, γ_1 and γ_2 in the GARCH(1,1) model.

The evolution from θ_t to θ_{t+1} is described in two steps: The GARCH equation (3) is the structural part, the evolution noise (4) adds a random disturbance. Positive definiteness of Σ_t is enforced by the functional form of the evolution equation. No restrictions on the GARCH parameters W_t, B_t and C_t are required.

Setting $B_t = 0$ would drop the GARCH component of the model, leaving a multivariate ARCH model. By setting $C_t = 0$ the model becomes a multivariate stochastic volatility model as discussed in Jacquier, Polson and Rossi (1995). With both, $B_t = 0$ and $C_t = 0$, the model would reduce to a random walk. The covariance matrix of the evolution noise was chosen using the concept of discount factors as $\Omega_t = \delta R_t$, where R_t is the posterior covariance matrix of the full parameter vector at time t. See West and Harrison (1989) for more discussion of the idea of using discount factors to specify evolution noise. We chose $\delta = 0.04$, implying an "information" loss of 4 percent through passage of one period of time.

4.2 *A Monte Carlo algorithm for the general dynamic model.* Let $D_t = D_{t-1} \cup \{y_t\}$ denote the information set at time t. Also, let $p(\theta_0|D_0)$ denote the original prior distribution, and let $p(\theta_t|D_t)$, $p(\theta_{t+1}|D_t)$, $t = 1, \ldots T$ denote the sequence of relevant posterior distributions in the general dynamic model, where $p(\theta_{t+1}|D_t)$, for example, reflects the information about the parameter vector at time $t+1$, after having observed $y_1, \ldots y_t$, but before seeing y_{t+1}.

In Müller (1992) a Monte Carlo integration scheme for general dynamic models was suggested which pushes a simulated Monte Carlo sample from the initial prior distribution through all the steps of the dynamic model, yielding for each step a Monte Carlo sample from the current posterior. The following steps describe the algorithm:

(a) The algorithm starts with a Monte Carlo sample $A_1 = \{\psi_1, \ldots \psi_m\}$, simulated from the original prior distribution: $\psi_i \sim p(\theta_1|D_0)$.

(b) Using a Metropolis type of algorithm this initial prior sample A_1 is transformed into a Monte Carlo sample $B_1 = \{\eta_1, \ldots \eta_m\}$ from $p(\theta_1|D_1)$. Details of this step are explained in the following paragraph.

(c) By direct simulation of the evoulution equation B_1 is transformed into a Monte Carlo sample $A_2 = \{\psi_1, \ldots \psi_m\}$ from $p(\theta_2|D_1)$. This is done by generating ω_i from the normal distribution given in the evolution equation (4) and setting $\psi_i = g(\eta_i) + \omega_i$, where g is the mapping implied by the evolution equation (3).

Continuing in an inductive way eventually provides a Monte Carlo sample from each of the relevant posteriors $p(\theta_t|D_t)$ and $p(\theta_t|D_{t-1})$. These Monte Carlo samples can then be used to estimate posterior integrals as desired.

Metropolis algorithm. Step (b) of the Monte Carlo scheme requires the transformation of the Monte Carlo sample $A_t = \{\psi_1, \ldots \psi_m\}$ from $p(\theta_t|D_{t-1})$ into a Monte Carlo sample B_t from $p(\theta_t|D_t)$. In the dynamic model context, $p(\theta_t|D_{t-1})$ can in general be expected not to be too far away from $p(\theta_t|D_t)$ because the additional observation y_t typically does not dramatically shift the posterior. This

feature makes application of the following algorithm attractive. Starting with a Monte Carlo sample point $x_0 = \psi_i$ from A_t simulate a Metropolis chain, $x_k, k = 0, 1, \ldots M$ with limiting distribution $p(\theta_t|D_t)$. Since the initial distribution is already close to the asymptotic distribution, only a few simulated iterations suffice to obtain an approximate draw x_M from $p(\theta_t|D_t)$.

Let $p(\cdot)$ denote $p(\theta_t|D_t)$. The Metropolis chain is formally described by the following steps:

(b.i) Start with $x_0 = \psi_i$.

(b.ii) Generate a "candidate" $\tilde{x} \sim h(y|x_0)$.

(b.iii) With probability $a(x_0, \tilde{x}) = \min(1, p(\tilde{x})/p(x_0))$ move to y, i.e., $x_1 := \tilde{x}$, else $x_1 = x_0$.

(b.iv) Repeat to generate x_2, x_3 etc. and stop after M iterations. Take x_M as approximate sample point from $p(\cdot) = p(\theta_t|D_t)$.

The "probing distribution" $h(\tilde{x}|x)$ in (b.ii) is chosen as $N(x, D)$ with D a diagonal matrix with elements $D_{ii} = c \cdot \sigma_i$, where σ_i is an estimate of the standard deviation of the i-th coordinate in $p(\theta_t|D_{t-1})$. The scaling factor c was chosen such that the acceptance probabilities are not too close to 1.0 or 0.0. After trying some values, $c = 0.10$ was found to be adequate. Too large steps (i.e., large c) would result in acceptance probabilities close to zero. Drawing candidates very close to the current point (i.e., small c) would lead to acceptance probabilities close to one. Neither is desirable. In the first case, the chain would get trapped in the current state, in the latter case the chain would take too long to move around the parameter space. Polson (1996) formalizes this argument in terms of the implied second eigenvalues of the Markov Chain.

A simplified mixture of Dirichlet process (MDP) model. The drawback of simulating the transition from $p(\theta_t|D_{t-1})$ to $p(\theta_t|D_t)$ by the Metropolis scheme is that it requires pointwise evaluation of the posterior $p(\theta_t|D_t)$ when evaluating the acceptance probabilities $a(x_k, \tilde{x})$. Analytic evaluation of $p(\theta_t|D_t)$ is practically impossible. Instead we use $p(\theta_t|D_t) \propto p(\theta_t|D_{t-1}) \cdot p(y_t|\theta_t)$, where we evaluate the likelihood $p(y_t|\theta_t)$ exactly, and substitute for $p(\theta_t|D_{t-1})$ a reconstruction $\hat{p}$ based on the available Monte Carlo sample A_t. We implement this reconstruction by an application of a mixture of Dirichlet process (MDP) density estimation model briefly described below. The use of the MDP model is not critical for the implementation of step (b). Many other alternative model based density estimation approaches would lead to very similar inference. For example, finite mixture of normal models could be used (Richardson and Green 1997).

The mixture of normal density estimation model described here is derived from the multivariate normal MDP model described in West, Müller and Escobar (1994). Let $\phi_{m,S}$ denote the $N(m, S)$ probability density function. Using the Monte Carlo sample $A_t = \{\psi_1, \ldots \psi_m\}$, the MDP model can be used to fit a mixture of normal kernels

$$\hat{p}(\theta_t) = \underbrace{\sum_{j=1}^{\infty} w_j \phi_{\mu_j,\Sigma}(\theta_t)}_{\int \phi_{\mu,\Sigma}(\theta_t) dG(\mu)}$$

to $p(\theta_t|D_{t-1})$. Let $\delta_x(\cdot)$ denote a pointmass at x. The estimate $\hat{p}(\cdot)$ can be thought of as a mixture of normal kernels $\phi_{\mu,\Sigma}(\cdot)$ with respect to a mixing measure $G(\mu) = \sum w_j \delta_{\mu_j}(\mu)$. Let ψ_i denote the elements of the Monte Carlo sample A_t. The MDP model defines a prior model for $\hat{p}$ by assuming that the unknown mixing measure G is generated from a Dirichlet process.

$$\psi_i \sim \int \phi_{\mu,\Sigma}(\psi_i) dG(\mu), \quad i = 1...m, \qquad \ldots(5)$$

$$\begin{aligned} G &\sim DP(G_0, \alpha), \\ G_0(\mu) &= N(a, B), \\ \Sigma^{-1} &\sim W(q, R^{-1}/q), \end{aligned}$$

Here $G \sim DP(G_0, \alpha)$ denotes that G is a random distribution generated from a Dirichlet process with base measure $\alpha \cdot G_0$. Also, $W(n, S)$ denotes a Wishart distribution with scalar parameter n and matrix parameter S.

The hyperparameters α, a, B, q and R are fixed at $\alpha = 1.0$, $a = \bar{\psi}$, $B = cov(\psi)$, $q = 41$, and $R = cov(\psi)$. Here $\bar{\psi}$ and $cov(\psi)$ are sample mean and covariance matrix of the Monte Carlo sample A_t. Estimation of the MDP model (5) is done by a Gibbs sampling scheme, described e.g. in West, Müller and Escobar (1994). The MDP model is an auxiliary model, used here only to obtain a reasonable approximation to $p(\theta_t|D_{t-1})$. Therefore, the Gibbs sampling scheme is run for just a few iterations, with the predictive density – conditional on the values of the MDP after those iterations – taken as the required approximation.

4.3 *Results.* We implemented the described Monte Carlo scheme to analyze the data set described in Section 2, using the three dimensional time series over the full time period, from November 23, 1988 until December 31, 1992. Let $diag(x)$ denote a diagonal matrix with the vector x as diagonal. As initial prior distribution $p(\theta_1|D_0)$ we chose independent normal distributions on all parameters with the following a priori means and standard deviations (in parentheses): $E(a_{1,j}) = 0$ (0.001), $j = 1,...3$, $E(A_{1,ij}) = 0$ (0.01), $i = 1,...3$, $j = 1,...3$, $E(B_{1,jj}) = E(C_{1,jj}) = 0.25$ (0.25), $j = 1,...3$, $E(B_{1,ij}) = E(C_{1,ij}) = 0$ (0.1), $i \neq j, i = 1,...3$, $j = 1,...3$, $E(W_{1,ij}) = E(V_{1,ij}) = L_{ij}$ (0.01). Here L is a Choleski decomposition of the sample covariance matrix over the period July 7, 1988 through November 22, 1988 (which were not included in the data). The particular choices formalize an a priori expectation of zero means and a covariance matrix corresponding to the previous 100 days, slightly inflated to allow for additional uncertainty. However, in the context of the dynamic model with evolution noise, the initial prior choice is not crucial since information is discounted

and the – over many periods – relatively strong information from the data soon dominates the initial prior.

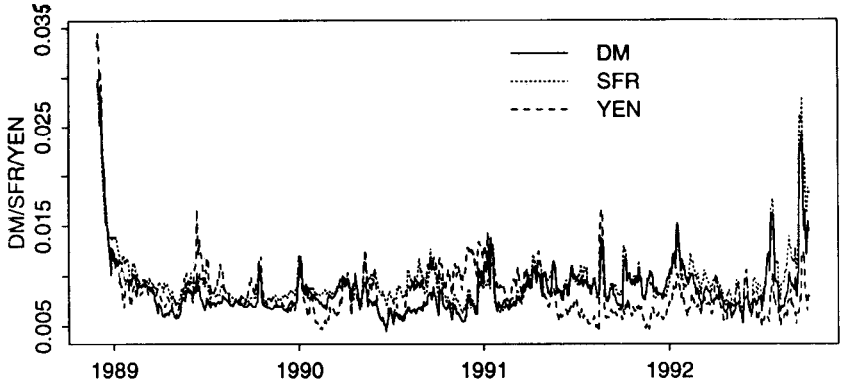


Figure 6: Posterior means $E(\sigma_{ii,t}|I_t)$ for mark, franc and yen. The initial burst is caused by the prior $p(\theta_1|D_0)$.

Figures 6 and 7 plot the trajectories of the one-step-ahead forecasts for standard deviations and correlations. While the graph seems to suggest strong fluctuations in the variances and correlations, this is misleading, at least on a daily basis. The graph represents the trajectory over more than 1000 days. When plotted over a shorter subinterval, the trajectories appear far smoother. Also, some of the big changes in level are due to the normality assumption in the observation equation (2). A single outlier in the data forces the variances to be adjusted upwards. Several authors (see, for example, Bollerslev *et al.*, 1992) noted that although the GARCH model explains some of the excess leptokurtosis observed in exchange rate data, it still falls short of explaining all. A possible variation of model (2) – (4) would consider multivariate t error distributions. Since the proposed implementation is based entirely on posterior simulation this would create no additional difficulties in the posterior simulation.

Figure 7 plots one-step ahead forecasts for correlations between the three series. Not surprisingly correlations between Deutsche Mark and French Franc are estimated uniformly higher than correlations between Mark and Yen and Mark and Swiss Franken. All three correlations are highly significant, confirming the importance of the joint analysis of all three time series in one multivariate model rather than three seperate univeriate analyses. Inference on correlations is critical from a perspective of decision problems related to model (2)–(4), including international portfolio management, option pricing etc. Conditional variances alone are poor proxies for risk (see, for example, Baillie and Bollerslev 1990).

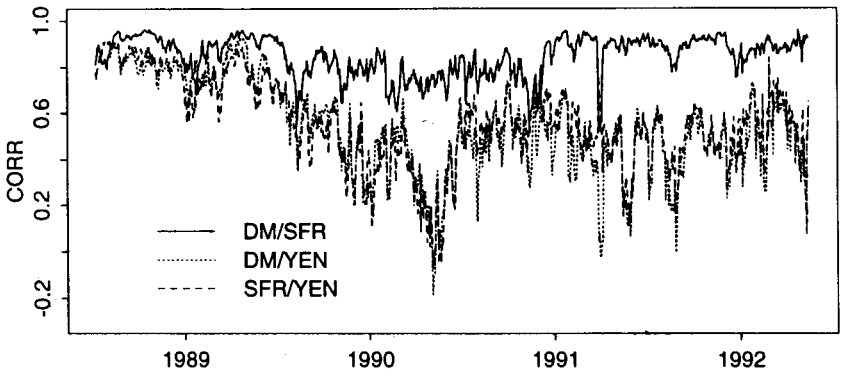


Figure 7: Posterior means for corr(mark,franc), corr(mark,yen) and corr(franc,yen).

The posterior means for the GARCH parameters B_t and C_t are shown in Figure 8. Figure 9 shows some elements of the posteriors $p(\theta_t|D_{t-1})$. The posterior densities shown in these plots are estimated by the MDP model (5). The density estimates show little deviation from multivariate normal contours, partially justifying the highly simplified version of the MDP model applied in this implementation. The near normal shape of the posteriors might even suggest moment updating and assuming normal distributions as an alternative to the density reconstruction. However, we would not recommend this in general, as parameter constraints, high evolution noise etc. could easily lead to a lack of posterior normality.

Figure 8 confirms the importance of the dynamic modelling framework. The posterior mean process for the parameters shows significant shifts over time. This is in agreement with our experience in analysing exchange rate data using GARCH models. Fitting GARCH models to seperate periods of the data yields substantially different parameter estimates. Treating the univariate series separately with individual univerate models reveals periods of sometimes great and sometimes small variance decay. Thus these simple analyses reveal the same time varying dynamics as evidenced in Figure 8.

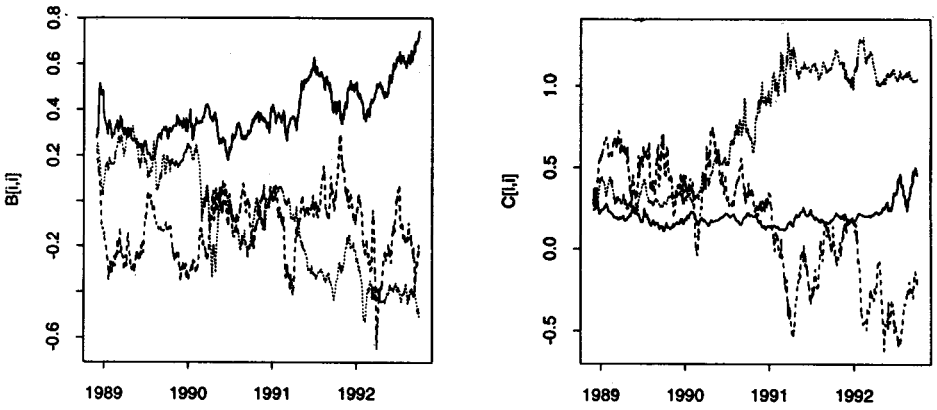
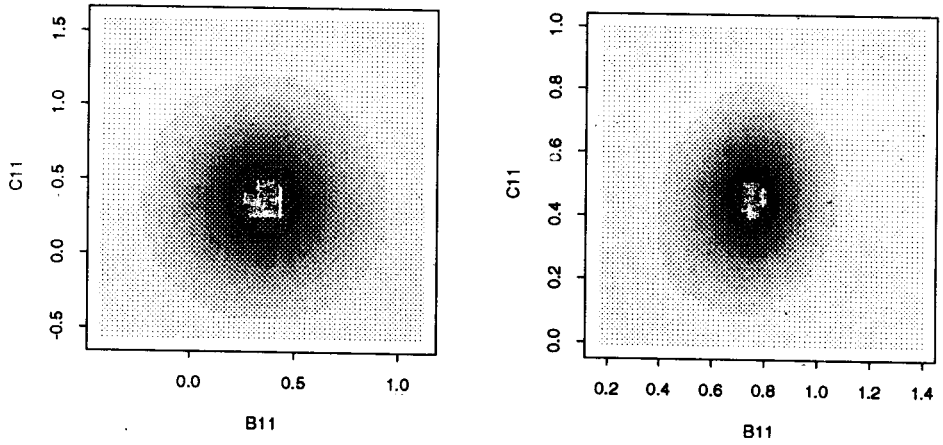


Figure 8: Posterior means for the diagonal elements of B_t and C_t ($B_{t,11}$,$C_{t,11}$ – solid line, $B_{t,22}$,$C_{t,22}$ – dashed line, $B_{t,33}$,$C_{t,33}$ – dotted line).



(a) At time $t = 7$.

(b) At time $t = 247$.

Figure 9: Posterior $p(B_{11,t}, C_{11,t}|D_{t-1})$ as reconstructed by the MDP model. The posterior at $t = 7$ still reflects the wide standard deviations of the initial prior. Because of the evolution noise in the model, also after 247 periods appreciable uncertainty about the parameters is left.

We conclude with some comments about our motivations for choosing the Bayesian paradigm in analysing the model. While we concede that our choice was predetermined by a fundamental preference for the Bayesian perspective (see, for example, Robert 1994), there are also other, more pragmatic arguments in this particular application. First, decision problems related to inference in exchange rate time series are typically focused on prediction and critically depend on a full description of all involved uncertainties. Both are natural in the context of Bayesian inference. Predictive inference is well defined as the posterior predictive distribution of, for example, future volatilities and correlations. The posterior distribution gives a full probabilistic description of all relevant uncertainties. Second, inference is possible by following standard posterior simulation schemes, albeit with some innovations, like the flexible posterior reconstruction using the mixture of normal model introduced in (5).

References

BAILLIE, R.T. and BOLLERSLEV, T. (1989). The message in daily exchange rates: A conditional variance tale. *Jour. Business and Economic Statistics*, **7**, 297–305.

BAILLIE, R.T. AND BOLLERSLEV, T. (1990). A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Jour. International Money and Finance*, **9**, 309–324.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Jour. Econometrics*, **31**, 307–327.

BOLLERSLEV, T., ENGLE, R.F., AND WOOLDRIDGE, M. (1988). A capital asset pricing model with time varying covariances. *Political Economy*, **96**, 116–131.

BOLLERSLEV, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, **72**, 498–505.

BOLLERSLEV, T., CHOU, R., AND KRONER, K.F. (1992). ARCH modeling in finance. *Jour. Econometrics*, **52**, 5–59.

CARLIN, B.P, POLSON, N.G. AND STOFFER, D.S. (1992). A Monte Carlo approach to non-normal and non-linear state-space modelling. *Jour. Amer. Statist. Assoc.*, **87**, 493-500.

ENGLE, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987-1008.

ENGLE, R.F. AND BOLLERSLEV, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, **5**, 1–50, 81–87.

ENGLE, R.F. AND GONZALEZ-RIVERA, G. (1991). Semiparametric ARCH models. *Jour. Business and Economic Statistics*, **9**, 345–360.

GALLANT, A.R. AND NYCHKA, W. (1987). Seminonparametric maximum likelihood estimation. *Econometrica*, **55**, 363–390.

GALLANT, A.R. AND TAUCHEN, G. (1989). Semi non-parametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica*, **57**, 1091–1120.

GELFAND, A.E. AND SMITH A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Jour. Amer. Statist. Assoc.*, **85**, 398-409.

GELMAN, A. AND RUBIN, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457 – 473.

GELMAN, A., ROBERTS, G. O. AND GILKS, W. R. (1996). Efficient metropolis jumping rules. In *Bayesian Statistics 5*, pp. 599-600, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford University Press.

GEWEKE, J. (1986). Exact inference in the inequality constrained normal linear regression model. *Jour. Applied Econometrics*, **1**, 127-141.

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57**, 1317-1339.

GEWEKE, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*, pp. 169-194, eds. J.O. Berger, J.M. Bernardo, A.P. Dawid and Smith, A.F.M, Oxford University Press.

HSIEH, D.A., (1989). Modeling heteroscedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistics*, **7**, 307–317.

– – – – (1988). The statistical properties of daily foreign exchange rates: 1974-1983. *Jour. International Economics*, **24**, 129-145.

JACQUIER, E., POLSON, N.G., AND ROSSI, P.E. (1994). Bayesian analysis of stochastic volatility models. *Jour. Business and Economics Statistics*, **12**, 371-389.

– – – – (1995). Models and priors for multivariate stochastic volatility. *Technical Report*, University of Chicago.

JORION, P. (1988). On jump processes in the foreign exchange and stock markets. *Review of Financial Studies*, **1**, 427–445.

LEE, T.K. (1988). Does conditional covariance or doncitional variance explain time varying risk premia in foreign exchange returns? *Ecomonics Letters*, **27**, 371–373.

MÜLLER, P. (1992). Posterior integration in dynamic models. *Computing Science and Statistics*, **24**, 318-324.

NELSON, D.B. (1990). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, **59**, 347–370.

POLE, A. and WEST, M. (1990). Efficient bayesian learning in nonlinear dynamic models. *Journal of Forecasting*, **9 (2)**, 119-136.

POLSON, N.G. (1996) Convergence of Markov Chain Monte Carlo algorithms. In *Bayesian Statistics 5*, pp. 292-322, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith, Oxford University Press.

RICHARDSON, S., and GREEN, P. (1997). On Bayesian analysis of mixtures with an unknwon number of components. *Jour. Royal Statist. Soci.*, Series B, **49**, 731-792.

ROBERT, C. (1994). *The Bayesian Choice*. Springer-Verlag, New York.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701-1728.

TSAY, R.S. (1987). Conditional heteroscedastic time series models. *Jour. Amer. Statist. Assoc.*, **82**, 590–604.

WEST, M. (1992). Mixture models, Monte Carlo, Bayesian updating and dynamic models. *Computing Science and Statistics*, **24**, 325-333.

WEST, M. AND HARRISON, P.J. (1989). *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.

WEST, M., MÜLLER, P., and ESCOBAR, M.D. (1994). Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of Uncertainty: A tribute to D. V. Lindley*, pp. 363-386, eds. A.F.M. Smith and P. Freeman.

WHISTLER, D. (1988). Semi-parametric ARCH estimation of intra daily exchange rate volatility. Unpublished manuscript, London School of Economics.

ZELLNER, A. and MIN, C-K. (1995). Gibbs sampler convergence criteria. *Jour. Amer. Statist. Assoc.*, **90**, 921–928.

PETER MÜLLER
INSTITUTE OF STATISTICS
AND DECISION SCIENCES
DUKE UNIVERSITY
DURHAM, NC 27708-0251
e-mail : pm@stat.duke.edu

ANDY POLE
INVICTUS PARTNERS
LLC, 590 MADISON AVENUE
37TH FLOOR, NEW YORK NY 10022.
e-mail : ap@stat.duke.edu