CORE DISCUSSION PAPER
9957

ADAPTIVE POLAR SAMPLING WITH AN APPLICATION TO A BAYES MEASURE OF VALUE-AT-RISK

Luc BAUWENS[1] Charles S. BOS[2] and Herman K. VAN DIJK[3]

October 1999

Abstract

Adaptive Polar Sampling (APS) is proposed as a Markov chain Monte Carlo method for Bayesian analysis of models with ill-behaved posterior distributions. In order to sample efficiently from such a distribution, location-scale transformation and a transformation to polar coordinates are used. After the transformation to polar coordinates, a Metropolis-Hastings algorithm is applied to sample directions and, conditionally on these, distances are generated by inverting the CDF. A sequential procedure is applied to update the location and scale.
Tested on a set of canonical models that feature near non-identifiability, strong correlation, and bimodality, APS compares favourably with the standard Metropolis-Hastings sampler in terms of parsimony and robustness. APS is applied within a Bayesian analysis of a GARCH-mixture model which is used for the evaluation of the Value-at-Risk of the return of the Dow Jones stock index.

Keywords: Markov chain Monte Carlo, simulation, polar coordinates, GARCH, ill-behaved posterior, Value-at-Risk.

JEL Classification: C11, C15, C63.

[1]CORE, Université Catholique de Louvain, Belgium. E-mail: bauwens@core.ucl.ac.be
[2]Econometric and Tinbergen Institutes, Erasmus University Rotterdam, P.O.Box 1738, NL-3000 DR Rotterdam, The Netherlands. E-mail: cbos@few.eur.nl
[3]Econometric Institute, Erasmus University Rotterdam. E-mail: hkvandijk@few.eur.nl

Correspondence to Charles S. Bos. We thank Michel Lubrano and participants at ESEM'99 for helpful comments on earlier versions of this paper. Support from HCM grant ERBCHRXCT 940514 is gratefully acknowledged.
This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

In recent years Markov chain Monte Carlo (MCMC) methods, in particular the Metropolis-Hastings (MH) and Gibbs samplers, have been applied extensively within Bayesian analyses of statistical and econometric models. The theory of Markov chain samplers dates back to Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) and Hastings (1970). A key technical reference on MCMC methods is Tierney (1994). Surveys oriented towards econometrics are provided by Chib and Greenberg (1996) and Geweke (1999). Applications are numerous, see e.g. Gelfand, Hills, Racine-Poon and Smith (1990), Albert and Chib (1993), Geweke (1993), Diebolt and Robert (1994), Bauwens and Lubrano (1998), Kim, Shephard and Chib (1998), Paap and van Dijk (1999), and Koop and van Dijk (1999).

Although MCMC methods revolutionized the applicability of Bayesian inference, there is, in practice, a substantial variation in their convergence behaviour. The sampling method or the structure of the model may be the culprit of such behaviour. Hobert and Casella (1996) show for instance that the Gibbs sampler does not converge when the posterior is improper. Justel and Peña (1996) emphasize the convergence problems of the Gibbs sampler when there are outliers. Kleibergen and van Dijk (1994,1998) indicate the near reducibility of MCMC methods when there exists near non-identifiability and non-stationarity in econometric models with flat priors. A common difficulty encountered using the MH sampler is the choice of a candidate density when little is known about the shape of the target density. In such a case, updating the candidate density sequentially is a partial solution.[1] The performance of the Gibbs sampler is seriously hampered by strong correlation in the target distribution. A multimodal target density may pose problems to both methods. If the MH candidate density is unimodal, with low probability of drawing candidate values in one of the two modes, this mode may be missed completely, even when the sample size is large. More generally stated, the acceptance probability may be very low, as many candidate values lying between the modes have to be rejected. With the Gibbs sampler, reducibility of the chain may occur in this case.

In this paper we introduce the method of adaptive polar sampling (APS) as an MCMC

[1] This corresponds to the experimental results obtained by local adaptive importance sampling when the posterior is ill behaved, see e.g. van Dijk and Kloek (1980), Oh and Berger (1992), and Givens and Raftery (1996).

algorithm to sample from a target (posterior) distribution which is possibly ill behaved. The algorithm features two transformations to induce a more regular shape of the target function in the transformed space than the one in the original space. The key transformation is one where the n-dimensional space is transformed into polar coordinates which consist of a distance measure and a $(n-1)$-dimensional vector of a direction (or angle). A MH algorithm is applied to sample the directions. Next, the distance is sampled conditionally on the direction by the inverse transformation method. A location-scale transformation is used prior to the transformation to polar coordinates and is sequentially updated, using the posterior first and second order moments obtained in successive rounds of the algorithm. The sequential procedure is intended to improve the acceptance rate of the MH step.

The advantages of the APS algorithm are twofold. Firstly, the algorithm is *parsimonious* in its use of information on the shape of the target density. Its location and scale need not be known precisely beforehand. Secondly, the algorithm is *robust*: It can handle a large variety of features of target distributions, in particular bimodality, strong correlation, extreme skewness, and heavy tails.

The outline of the paper is as follows. In Section 2 the algorithm is introduced. In Section 3 four canonical models are described. These models are used for experimenting with APS and for comparing its performance with that of the standard Metropolis-Hastings algorithm. The models include a normal distribution, for testing the behaviour of the algorithm on a well-behaved target, an ARMA-GARCH model with strong correlation between the parameters, a bivariate mixture with bimodality, and an ARCH-variance mixture where the parameters may be badly identified.

We also apply APS to an empirical example. In Section 4 we investigate the daily returns on the US stock market. The data is modelled using a GARCH-mixture model. The object of the analysis is to investigate the effect on the Value-at-Risk of the mixture component in the model for the returns on the stock index. Conclusions are presented in Section 5.

2 Adaptive polar sampling

MCMC methods usually generate random drawings in the original parameter space. If the target density is not well behaved, the chosen algorithm may converge very slowly, or not at all. The method presented here transforms the parameter space into a $(n-1)$-dimensional

space in which the density is reasonably well behaved, and a unidimensional complementary space in which most of the ill behaviour is concentrated.

The method is first explained using an example. In panel **A** of Figure 1 a bivariate mixture density of a parameter vector $\theta = (\theta_1\, \theta_2)$ is shown, which incorporates the problems of strong correlation between θ_1 and θ_2 (the correlation is -0.84) and displays the bimodality. A location-scale transformation changes the density such that the resulting density in the $(y_1\, y_2)$ space has its location at the origin and a unit covariance matrix (see panel **B**). This transformation spreads the mass of the density around the origin as much as possible.

The second transformation changes the $(y_1\, y_2)$ coordinates to polar coordinates with direction (or angle) η and distance ρ from the origin.[2] Panel **C** depicts the (u-shaped) marginal density $p(\eta)$ of the direction, with the dotted line showing the (uniform) density of η in the case the original density would have been a bivariate normal. We sample η by comparing its marginal density to the one resulting from a normal density after the same transformation, using a Metropolis-Hastings step. Most of the ill behaviour of the density in the original parameter space is concentrated in the conditional density of ρ given a value of η. Panels **D** and **E** illustrate the difference between the density of $\rho|\eta$ resulting from the normal density (the dotted line) and from the mixture density in Panel **A** (continuous line). In Panel **D** the conditional density for $\eta = 0$ is drawn; in Panel **E** an angle of $\eta = \frac{1}{4}\pi$ is chosen. In both cases it is found that the density resulting from the normal density is not a good approximation for the target densities.

Defining the transformations

The transformation is indicated by

$$(\eta, \rho) = T(\theta \,|\, \mu, \Sigma) = T_{y\to\eta,\rho}(y) \circ T_{\theta\to y}(\theta \,|\, \mu, \Sigma), \tag{1}$$

where θ is the original parameter, y is the standardized parameter, and (η, ρ) are the polar coordinates. One starts by standardizing to a variable y, conditional on estimates of location μ and scale Σ,

$$y = \Sigma^{\frac{1}{2}}(\theta - \mu), \tag{2}$$

[2]The distance measure we use is a signed measure, which explains why ρ can take negative values in Panels **D** and **E** of Figure 1. See below for details.

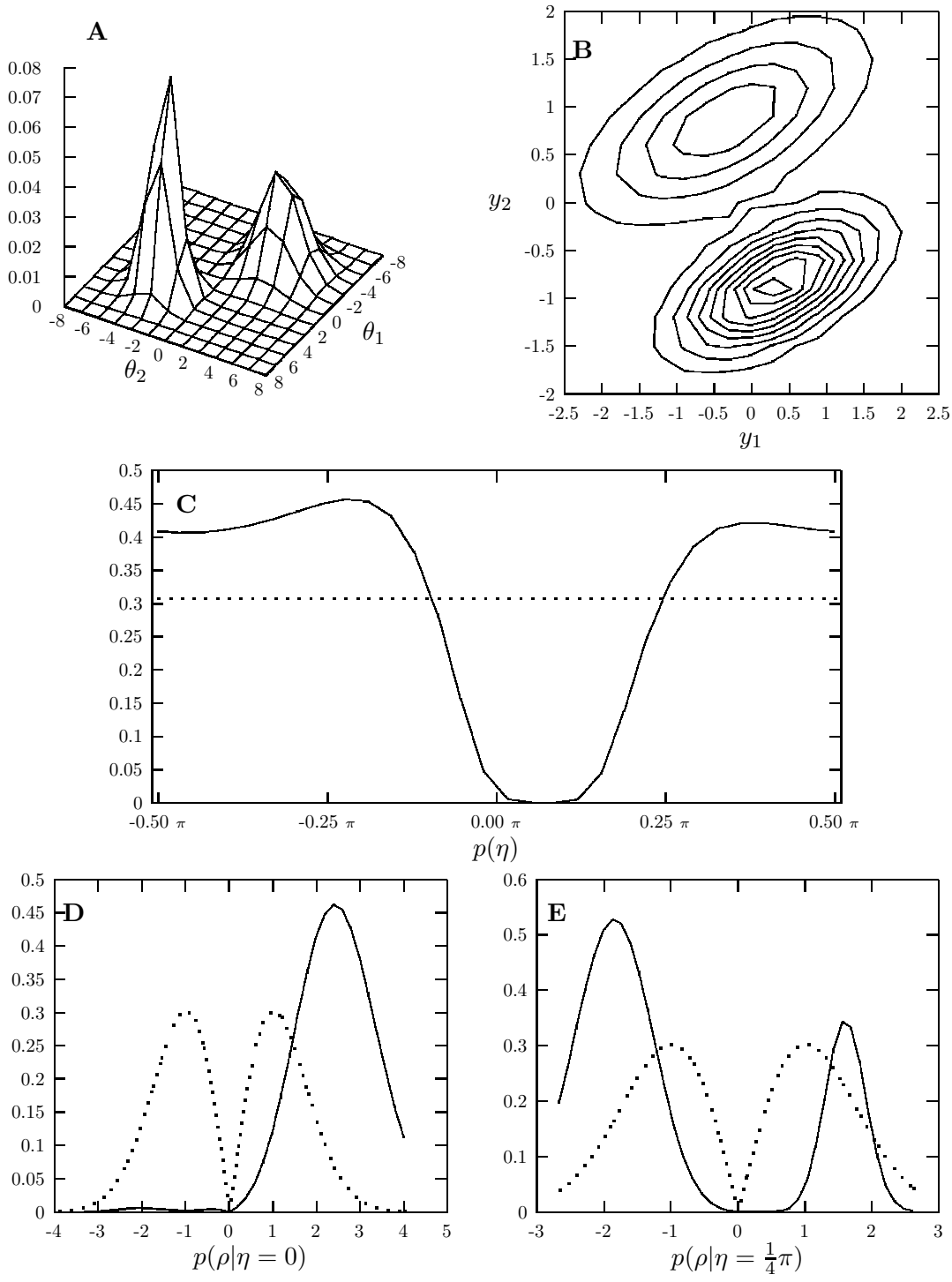

Figure 1: Transforming a bivariate mixture density.

with $\Sigma^{\frac{1}{2}}$ the Cholesky decomposition of Σ. Next, define d as the length of the vector y,

$$d = \sqrt{y'y}. \tag{3}$$

The signed distance measure ρ is based on d, but also indicates the sign of the first element of y:

$$\rho = \text{sgn}(y_1)\, d. \tag{4}$$

The directions η_j are defined recursively by

$$\eta_j = \arcsin \frac{y_{n-j+1}}{\rho \prod_{i=1}^{j-1} \cos \eta_i} \in (-\frac{1}{2}\pi, \frac{1}{2}\pi] \qquad \text{for } j = 1, .., n-1, \tag{5}$$

where by convention $\prod_{i=1}^{0} \cos \eta_i = 1$, so that $\eta_1 = \arcsin y_n/\rho$. The inverse transformation $T^{-1}_{y \to \eta,\rho}(\eta, \rho)$ from polar coordinates to y, is

$$y_j = T^{-1}_{y,j}(\eta, \rho) = \begin{cases} \rho \sin \eta_{n-j+1} \prod_{i=1}^{n-j} \cos \eta_i & \text{if } j = 2, .., n \\ \rho \prod_{i=1}^{n-1} \cos \eta_i & \text{if } j = 1. \end{cases} \tag{6}$$

The Jacobian of the transformation $T_{\eta,\rho}(\theta)$ from θ to (η, ρ) is

$$\begin{aligned} J(\eta, \rho) &= \det(\Sigma)^{\frac{1}{2}} \times |\rho|^{n-1} \times \left| \prod_{i=1}^{n-2} \cos^{n-i-1} \eta_i \right| \\ &= \det(\Sigma)^{\frac{1}{2}} \times |J(\rho)| \times |J(\eta)|. \end{aligned} \tag{7}$$

The difference between the transformation in (4)-(5) and the usual transformation to polar coordinates—see e.g. Kendall and Stuart (1973), p. 247—stems from the inclusion of the sign function in the distance measure.[3] This inclusion leads to different domains for the angles and for the distance. In the usual case, the domain of η is $\Omega_\eta = [0, 2\pi) \times (-\frac{1}{2}\pi, \frac{1}{2}\pi]^{n-2}$, whereas in our case $\Omega_\eta = (-\frac{1}{2}\pi, \frac{1}{2}\pi]^{n-1}$.[4] The advantage of our transformation stems from the way the sampler is set up (see below for details): in the space of η it is important to have a smooth, well-behaved density, whereas it does not hamper the algorithm if ill behaviour is

[3] In the sequel, we refer to ρ as a distance even though it can be negative.

[4] In the bivariate case, with the usual transformation, a point in R^2 is described by an angle between 0 and 2π (a full circle)—which defines a direction (a straight line through the origin)—and the Euclidean distance from the point to the origin. In our version, the same point is described by an angle between $-\pi/2$ and $\pi/2$ (a half circle) and the signed Euclidean distance. The sign of the distance identifies on which side of the y_2 axis the point is located.

concentrated in the density of ρ. A smaller range of η may help in this respect. Note that the density of η in our transformation corresponds to a weighted average of the density at locations η and $\eta + \pi$ in the usual transformation.

Remark: The transformation used here is similar to the transformation used in the mixed integration algorithm of van Dijk, Kloek and Boender (1985). The mixed integration method combines importance sampling on the directions with numerical integration on the distance. It is an algorithm for computing posterior moments of functions of parameters. As such, it does not deliver draws of the posterior density, contrary to APS.

Sampling directions and distances

Our proposal for sampling the directions is to use a standard MH algorithm. Consider the case where APS reaches a point $\theta^{(i)} = T^{-1}(\eta^{(i)}, \rho^{(i)})$, where the subscript i denotes the $i-$th drawing. Note that we suppress the dependence of the transformation in Equation (1) on μ and Σ for notational convenience. A candidate θ^* is drawn from a normal distribution with expectation μ and covariance matrix Σ. The new direction η^*, calculated using the transformation[5] $(\eta^*, \cdot) = T(\theta^*)$ does not depend on the previous $\theta^{(i)}$ or $\eta^{(i)}$. That is, we use the independent MH sampler. The acceptance probability of a move from $\eta^{(i)}$ to η^* is given by

$$\alpha(\eta^{(i)}, \eta^*) = \min \left[\frac{p(\eta^*) q(\eta^{(i)})}{p(\eta^{(i)}) q(\eta^*)}, 1 \right], \tag{8}$$

where $p(.)$ and $q(.)$ denote the marginal target and candidate densities of η. If the drawing is accepted, then $\eta^{(i+1)} = \eta^*$, else the algorithm continues with $\eta^{(i+1)} = \eta^{(i)}$.

Using a normal candidate $\theta \sim N(\mu, \Sigma)$ implies that $q(\eta) \propto |J(\eta)|$ as defined by (7) (see the Appendix). The marginal target density is defined by transforming the target density $p_\theta(\theta)$ through polar coordinates and marginalizing with respect to ρ:

$$p(\eta) = \int_{-\infty}^{\infty} p_\theta(T^{-1}(\eta, \rho)) \, |J(\eta)| \, |J(\rho)||J(\Sigma)| \, d\rho. \tag{9}$$

Hence, the ratio of densities in (8) (where for notational convenience the superscript (i) is

[5] In the bivariate case, it is easy to draw a candidate η^* directly from its uniform distribution. However, in the multivariate case ($n > 2$), the distribution of η is not standard, therefore we draw from the normal density and apply the transformation.

deleted) can be simplified as follows:

$$\frac{p(\eta^*)q(\eta)}{p(\eta)q(\eta^*)} = \frac{|J(\eta^*)|\,|J(\Sigma)|\int p_\theta\left(T^{-1}(\eta^*,\rho)\right)|J(\rho)|\,\partial\rho}{|J(\eta)|\,|J(\Sigma)|\int p_\theta\left(T^{-1}(\eta,\rho)\right)|J(\rho)|\,\partial\rho}\frac{|J(\eta)|}{|J(\eta^*)|}$$
$$= \frac{\int p_\theta\left(T^{-1}(\eta^*,\rho)\right)|J(\rho)|\,\partial\rho}{\int p_\theta\left(T^{-1}(\eta,\rho)\right)|J(\rho)|\,\partial\rho}. \tag{10}$$

We note that such a simplification is available for all densities which can be written as $q(\theta) = f\left((\theta-\mu)'\Sigma^{-1}(\theta-\mu)\right)$.

The marginal target density of η defined by Equation (9) can be calculated efficiently using a unidimensional numerical integration method.

The sampling is completed by drawing $\rho^{(i+1)} \sim p(\rho\,|\,\eta^{(i+1)})$. The information accumulated during the numerical integration for calculating $p(\eta^*)$ can be used in a straightforward way for the construction of a numerical approximation to the conditional density of ρ. Based on the distribution function the inverted CDF method provides a drawing $\rho^{(i+1)}|\eta^{(i+1)}$. Next, $\theta^{(i+1)} = T^{-1}(\eta^{(i+1)}, \rho^{(i+1)})$ is computed, and a new draw of θ can be generated.

Note that in the bivariate case APS is similar to the Box-Muller method for generating draws of the standard normal density. The Box-Muller method relies on the usual polar coordinates. In practice, it is implemented by drawing independently two uniform variates in (0,1) and transforming these. See Rubinstein (1981), p. 86. This method can be interpreted as sampling an η from a uniform density on $[0, 2\pi)$ and using the inverse transformation method for generating a $\rho^2 \sim \exp(1)$. By transforming the polar coordinates to standard coordinates, two independent drawings from a standard normal density result. In APS the uniform density of η is used as an approximation to the target density of η, and the target density of ρ is inverted numerically since it is not available analytically.

Updating location and scale

The transformation to polar coordinates depends on an estimate of μ and Σ of the $N(\mu, \Sigma)$ candidate. If these estimates correspond reasonably to the posterior first and second order moments then the distribution of the directions implied by the normal candidate density should mimic that of the directions of the posterior distribution. This results in reasonable acceptance rates in the MH step of APS. If the estimates of the mean and covariance matrix of the target density are not accurate, sequential updating of μ and Σ may be applied. The

updating may continue until the acceptance rate is high enough (or no longer increases). In the final round, a large sample of θ may be generated.

A summary of the steps of one iteration of the algorithm is presented in Table 1.

Table 1: APS algorithm (summary)

1. Generate θ from $N(\mu, \Sigma)$
2. Transform θ to $y = \Sigma^{\frac{1}{2}}(\theta - \mu)$
3. Transform y to η by (4)-(5)
4. Apply MH step on η
5. Generate $\rho\|\eta$ by inverting numerically its CDF
6. Transform (η, ρ) to y by (6)
7. Transform y to $\theta = \mu + \Sigma^{-\frac{1}{2}} y$

Convergence

APS is a combination of a Metropolis-Hastings sampler on the directions η and an inverted CDF method for generating the distance ρ. The MH step on the directions introduces dependence on past drawings. Hence, the transition kernel of APS is just the transition kernel of the MH step. It can be written as

$$K(\eta^{(i)}, \eta^{(i+1)}) = \begin{cases} q(\eta^{(i+1)})\, \alpha(\eta^{(i)}, \eta^{(i+1)}) & \text{if } \eta^{(i+1)} \neq \eta^{(i)} \\ 1 - \int_{\eta \in \Omega_\eta} q(\eta) \alpha(\eta^{(i)}, \eta) d\eta & \text{if } \eta^{(i+1)} = \eta^{(i)}, \end{cases} \tag{11}$$

see e.g. Geweke (1999). The following theorem, adapted from Chib and Greenberg (1995), provides sufficient conditions for convergence of the adaptive polar sampler.

Theorem 2.1 *If for every $\eta \in \Omega_\eta$, the density $p(\eta)$ is non-null, and if for all combinations $(\eta^{(i)}, \eta^{(i+1)}) \in \Omega_\eta \times \Omega_\eta$, the densities $p(\eta^{(i)})$ and $q(\eta^{(i+1)})$ are positive and continuous, then the APS kernel defined by* (11) *is ergodic, and the sampled chain of random drawings converges in distribution to the target distribution.*

The first condition implies that some probability mass lies in all directions $\eta \in \Omega_\eta = (-\pi/2, \pi/2]^{n-1}$. This condition is fulfilled when μ lies in the interior of the original parameter space. The second condition holds for $q(\eta)$ since $q(\eta) \propto |J(\eta)|$, which is positive and continuous. The target density $p(\eta)$ is usually positive for all η as discussed above. Its continuity depends on the target function in the original space. If the original target density is not

continuous, consisting for example in blocks of mass separated by regions where $p_\theta(\theta) = 0$, the shape of the density $p(\eta)$ should be inspected carefully.

The APS algorithm can be expected to behave at least as good as the MH sampler. This can be understood by noticing that APS approximates the posterior density exactly for one parameter (ρ) and uses a MH step for the remaining parameters (in the vector η). Given that η is a subvector of a one-to-one transformation of y (see Equations (4)-(6)), the ratio of densities of η, see (8), is the same whether one works in the space of η or in the space of y. The ratio of densities in the y-space for the case of the APS algorithm is, however, part of the ratio of densities in the standard MH algorithm. It can be shown that the building blocks of the acceptance probability for the n-dimensional MH algorithm factorise into a part common with the APS algorithm, and a part specific to the comparison of densities of distances ρ in the previous and candidate (new) direction. This latter part does not occur in the acceptance probability of the APS sampler. The exact gain in acceptance probability depends on the shape of the posterior. A set of experiments is reported in the next section in order to illustrate the improvement in four practical cases.

3 Comparison of methods

The APS algorithm is proposed as a parsimonious and robust sampling scheme. In this section, we test APS experimentally on four canonical models and compare its performance with a standard MH algorithm.

Models

The set of target distributions to be sampled from is as follows:

A: A six-dimensional normal density, with mean $\mu = (1, 2, , 4, 5, 6)'$ and covariance matrix $\Sigma = \text{diag}(6, 5, 4, 3, 2, 1)$.

B: The likelihood of an ARMA-GARCH model with near-cancellation of the parameters

of the ARMA part. A dataset of 200 datapoints is generated from the model

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \\ \epsilon_t &\sim \mathcal{N}(0, h_t) \\ h_t - \delta h_{t-1} &= \omega + \alpha \epsilon_{t-1}^2 \end{aligned}$$

with parameters chosen as $\phi = 0.5, \theta = -0.3, \delta = 0.6, \omega = 0.1$ and $\alpha = 0.1$. The target density is the likelihood function of the parameters. The expected large correlation between ϕ and θ would slow convergence of a Gibbs algorithm.

C: A bivariate mixture of normals. The target distribution is

$$\pi(\theta) = (1-p)\mathcal{N}(d\mu, \Sigma_1) + p\mathcal{N}(-d\mu, \Sigma_2)$$

The parameters are fixed at $p = 0.5, \mu = (-1, 1)', d = 8, \Sigma_1 = I_2$ and $\Sigma_2 = 2\Sigma_1$.

This is a case for which the Gibbs sampler leads to a (nearly) reducible chain, as the probability of switching from one mode to the other is very low. Also, the MH algorithm should be tuned carefully to ensure that this sampler does not miss one of the modes completely. Problems with convergence can be expected to increase with the value of d.

D: The posterior distribution of the parameters in an ARCH-mixture model. The model used for generating a dataset is

$$\begin{aligned} y_t &= \sqrt{h_t} u_t \\ h_t &= \omega + \alpha y_{t-1}^2 \\ u_t &\sim (1-p)\mathcal{N}(0,1) + p\mathcal{N}(0, (1+c)^2) \end{aligned}$$

The target distribution is the posterior of the parameters $\theta = (\omega, \alpha, p, c)$, where the prior is

$$\pi(\omega, \alpha, p, c) = \frac{1-\alpha}{\omega} \times I_{0 \le \alpha < 1} \times I_{\omega > 0} \times \frac{1}{(1-p) + p(1+c)^2} \times I_{c>0} \times I_{0.05 < p < 0.95}$$

This prior is based on the idea of the non-informative prior for the variance components in the ARCH and mixture parts; see Section 4 for details.

For generating the data the parameters where fixed at $\omega = 1, \alpha = 0.6, p = 0.2$ and $c = 1$, so that the u_t error process has a fourfold larger variance in 20% of the cases. Again, 200 datapoints were generated from the model.

Implementation issues

For implementing the experiments some choices have to be made concerning initial values, number of drawings per round, number of rounds, and diagnostics for assessing convergence. For APS, four rounds of updating are allowed for improving the location and scale of the candidate density, in order to increase the acceptance rate. The number of different directions η that is drawn, differs over the rounds. After drawing only 100 directions in the first two rounds, 500 directions are generated in the third round and 1000 in the last. In each round a sample of 10000 θ is drawn in newly accepted directions, thus the number of θ in a single direction goes down from 100 in the first round to 10 in the last round. In the first round, the MH step is skipped, to get a quick improvement over the initial estimate for the location and scale. In the last round, the initial 1000 θ are disregarded, to allow for a burn-in period.

For comparability with APS, the standard Metropolis-Hastings sampler uses an independent normal density for generating candidate draws. Initially, the scale is chosen such that the candidate density is very much spread out, which lowers the acceptance rates in the first rounds, but improves the probability of not missing a mode (as in the example of the bivariate mixture distribution). We run the algorithm over four rounds, with the last one containing 10000 accepted θ to be comparable to the sample of APS. Previous rounds contain less θ. We take successively 1000, 1000, 5000 and 10000 accepted θ.

Each sampler used the same dataset for the same model. The initial condition that was used in all samplers was fixed at a point far away from the mode(s) of the target. The scale matrix was fixed at 10 times the identity matrix, to search for the density function initially over a large region.

For the numerical integration in the APS algorithm, we used a 12 point adaptive Gauss-Legendre technique. Moreover, the parameter space was bounded in such a way that practically all probability mass lies within bounds. Note that theoretically there is no need to bound the parameter space, only for matters of computational efficiency it is convenient to limit the range where the probability mass is searched for.

Sampling results

Several diagnostics have been proposed to assess the convergence of Markov chains and the precision with which certain aspects of the distribution can be estimated. Some important

Table 2: Results for the Adaptive Polar Sampling algorithm

	Normal	**Arma-GARCH**	**Normal mixture**	**Arch mixture**
Acc. rate 2	0.06	0.10	0.37	0.23
Acc. rate 3	0.83	0.58	0.70	0.71
Acc. rate 4	0.89	0.72	0.71	0.73
Mahalanobis 2	5.42	9.52	0.05	4.00
Mahalanobis 3	0.17	0.20	0.00	0.44
Mahalanobis 4	0.00	0.01	0.00	0.08
Autocorrelation ρ_1^*	-0.06	0.51	0.09	0.39
Fn. ev. (last round, $\times 1000$)	110	101	175	103
Fn. ev. (total, $\times 1000$)	337	252	293	149

references on the subject are Geweke (1992), Roberts and Smith (1994), and Chib and Greenberg (1996). In this paper, we make use of the acceptance rate, the Mahalanobis distance,[6] and the largest (over parameters) first order autocorrelation coefficient between drawings.

Tables 2 and 3 present the results for the APS algorithm and for the standard MH algorithm, respectively. The first block in each table provides the acceptance rates of each round, after a preliminary start-up round. For APS, the acceptance rate refers to the directions, and for the standard MH method, it refers to the θ drawings. In case of the normal and the ARMA-GARCH models, the acceptance rate of APS in the second round is quite low, indicating that the improvement over the initial location and scale parameters (which were fixed at some extreme, incorrect values for all models) is not on target yet. After the third round, however, for all models acceptance rates of at least 70% are found. This is a good score, especially in comparison with the acceptance rates obtained with the standard MH algorithm. Only for the normal model (where the candidate density used equals the target density, when the location and scale are estimated perfectly), the acceptance rate of the standard MH algorithm is better, for the other models rates between 25 and 54 percentage points lower. Notice that in the case of the ARMA-GARCH model, convergence of the standard MH method seems to be slow, judging from the acceptance rates that do not rise quickly over the different rounds. The acceptance rate for the bivariate mixture model using the MH sampler depends strongly on the distance between the modes. The results in Tables 2 and 3 are gen-

[6]Defined as $(\mu^{[j]} - \mu^{[j-1]})'(\Sigma^{[j]})^{-1}(\mu^{[j]} - \mu^{[j-1]})$ where the superscript j indicates the j-th round of the algorithm.

Table 3: Results for the Metropolis-Hastings algorithm

	Normal	**Arma-GARCH**	**Normal mixture**	**Arch mixture**
Acc. rate 2	0.31	0.03	0.05	0.36
Acc. rate 3	0.86	0.08	0.14	0.52
Acc. rate 4	0.96	0.45	0.17	0.49
Mahalanobis 2	1.24	0.65	1.69	0.07
Mahalanobis 3	0.02	0.61	0.44	0.05
Mahalanobis 4	0.00	0.08	0.00	0.05
Autocorrelation ρ_1^*	0.08	0.81	0.83	0.63
Fn. ev. (last round, $\times$1000)	11	24	66	12
Fn. ev. (total, $\times$1000)	53	161	158	28

Table 4: Results for the normal mixture model, varying distances

d	0	2	4	6	8	10
APS	0.99	0.82	0.72	0.71	0.71	0.71
MH	0.95	0.57	0.31	0.21	0.17	0.98

The MH Sampler did not converge to the correct location/scale estimates when the mixture with $d = 10$ was sampled.

erated using a distance parameter $d = 8$ in model **C**. Table 4 reports the acceptance rates of both algorithms for a series of values of the distance parameter. The acceptance rate of the APS algorithm hardly changes between the different settings, whereas the final behaviour of the MH algorithm depends strongly on the gap between the modes. When sampling the bivariate mixture with modes lying far apart, the risk is high that the MH sampler does not sample from both modes. Indeed, the entry in the last cell of Table 4 indicates that the mean and location got to be adapted to just one, normal, mode, from which sampling could be done efficiently. The high acceptance rate for the MH sampler is in this case an indication of non-convergence, not of efficient sampling.

In APS, after two rounds, the algorithm has adapted the location parameter to a large extent, see the second block of Table 2. As the number of draws of θ is set up to be comparable between the APS algorithm and the MH algorithm, we expect similar behaviour of the Mahalanobis distance upon convergence.

The autocorrelation coefficient for the MH algorithm is higher for all models. This could be expected, as the Markov chain for the MH algorithm relates to the θ directly, whereas in

the case of APS the relation is through the directions η. Therefore, dependence in the APS-Markov chain of θ's is essentially of a nonlinear nature, leading to less (linear) correlation between sampled parameter vectors.

The bottom block of the tables reports the number of function evaluations, both in the last round and in total. Note that in the last round of APS more directions are evaluated (numerically), leading to a relatively high number of function evaluations. The MH algorithm uses less function evaluations per drawing of θ, as no numerical unidimensional distributions have to be calculated. However, this comes at the cost of a slower convergence, depending on the shape of the target distribution.

Summarizing, it is found that the APS algorithm is flexible and robust, in the sense that it can be used on a range of different target distributions, and it is parsimonious in its use of prior information on the shape of the target. Its efficiency depends on the model used. The number of function evaluations is larger than in the case where the standard MH sampler is used. This is the price to pay for the extra robustness. However, the final sample of APS is of better quality than that of the standard MH algorithm, as APS exhibits less autocorrelation in the chain, and the acceptance rate is usually higher than in the case of standard MH sampling.

4 Value-at-Risk analysis of the returns on the Dow Jones Stock Index

Data description

Daily returns on the stock market indices are series which typically exhibit low or nonexistent autocorrelation. The volatility however is known to cluster. A basic approach to modelling returns on stock markets is to use GARCH models; see e.g. the papers collected in Engle (1995).[7]

In our application we use daily data on the Dow Jones stock index from January 1, 1992 until November 17, 1998. The index is transformed into daily returns by taking the first differences of logarithms and multiplying by 100. This resulted in a dataset of 1794 observations that are displayed in Figure 2.

[7] Another approach is to use a stochastic volatility model, see e.g. Jacquier, Polson and Rossi (1994).

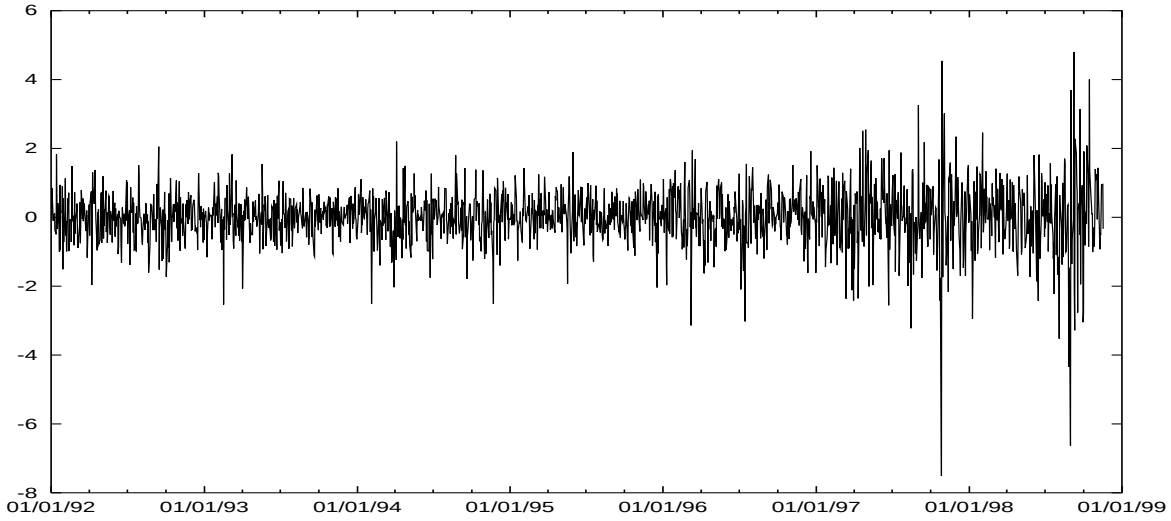

Figure 2: Returns on Dow Jones index

Table 5: Descriptive Statistics on Returns of Dow Jones index

Part of data	100%	99%	95%
Sample size	1794	1776	1704
Mean	0.0581	0.0616	0.0681
St. deviation	0.8296	0.7265	0.6152
Skewness	−0.6935	−0.2458	−0.0855
Kurtosis	11.6847	4.0939	2.8646
Autocorrelation	0.0095	0.0298	0.0168

Table 5 reports a set of descriptive statistics from the complete sample and from trimmed samples after deleting 1% and 5% of the extreme (both high and low) observations. For the full sample it is seen that the average daily return in this period was around 0.06 percent, with a relatively large standard deviation. The first-order autocorrelation is so small that it may be neglected in the modelling of this dataset. Looking at the graph of the returns, the alternation between tranquil and more volatile periods is apparent. Also, one can spot several outliers in the data, with more large negative outliers occurring than positive ones. The data exhibit thick tails, judging from the high kurtosis that is found. The statistics for the trimmed samples indicate that a small proportion of outlying observations are very influential on these statistics (except the mean): Skewness is largely removed by deleting a few outliers, just like the excess kurtosis, which goes down from 11.7 to around 4, after deleting only 1% of the data.

The heteroskedasticity of the returns can be modelled by a GARCH process. Since skewness is related to a few outlying observations, it may be sufficient to allow for a small percentage of aberrant data points. A variance mixture is useful in this respect. By allowing a (small) fraction of the disturbances to come from a higher-variance distribution, extreme observations are allowed for.

Model specification

The GARCH-mixture model is the GARCH version of model **D** in Section 3, and is written as

$$
\begin{aligned}
y_t - \overline{y} &= \epsilon_t \\
\epsilon_t &= \sqrt{h_t} u_t \\
h_t &= \omega + \alpha \epsilon_{t-1}^2 + \delta h_{t-1} \\
u_t &\sim (1-p)\mathcal{N}(0,1) + p\mathcal{N}(0,(1+c)^2) \\
\delta &> 0, \omega \geq 0, \alpha > 0, c \geq 0, 0 < p < 1
\end{aligned}
$$

with $y_t - \overline{y}$ the return in deviation from its mean. A fraction $1-p$ of the disturbances u_t follows the standard normal distribution; a small fraction p has a standard deviation which is increased to $1+c$. The restrictions on the parameters are the usual ones for identifiability and positivity. This GARCH-mixture model simplifies to the simple GARCH model if c or p

are fixed at zero.

Prior densities

In GARCH models, it is hard to devise reasonable informative priors directly on the parameters. Instead, we devise a prior which describes information in the space of the unconditional variance of the simple GARCH process; see Nelson (1990) and Kleibergen and van Dijk (1993) for a discussion of stationarity in GARCH models. This variance is

$$\sigma^2_{\text{GARCH}} = \frac{\omega}{1-\delta-\alpha}. \tag{12}$$

We choose the prior on the GARCH parameters to be non-informative on this unconditional variance, which gives

$$\pi(\delta, \omega, \alpha) \propto \frac{1-\delta-\alpha}{\omega} \times I_{\delta+\alpha<1} \times I_{\omega>0}. \tag{13}$$

The restriction on the sum of δ and α is sufficient for the simple GARCH process to be covariance stationary.[8]

If the fraction of disturbances u_t stemming from the distribution with the high variance is too close to one of the bounds, identifiability of the parameters can be problematic. Therefore, we limit a priori p to lie within the range $(0.01, 0.3)$. Within this range, we assume a uniform prior on p.

The parameters c and p determine the variance of the disturbance process u_t, which is

$$\sigma^2_u = (1-p) + p(1+c)^2. \tag{14}$$

The combined prior of p and c is non-informative on this variance, i.e.

$$\pi(p, c) \propto \frac{1}{(1-p) + p(1+c)^2} \times I_{0.01<p<0.3}. \tag{15}$$

The joint prior on all the parameters is the product of the priors in (13) and (15).

[8] We could also define the prior to be proportional to the inverse of the unconditional variance of the GARCH-mixture model. The difference would be that in the right hand side of (13), α would be multiplied by $\text{var}(u_t)$ as given by (14). We refrain from using this prior in order to impose prior independence between the GARCH parameters and the mixture parameters.

Posterior results

The first five panels in Figure 3 depict the marginal posterior densities of the parameters $\delta, \omega, \alpha, p$, and c. The corresponding posterior means and standard deviations are reported in the first block of Table 6.

The mode of the posterior density of the GARCH parameter δ lies around 0.93, which indicates that on a day to day basis the variance of the returns is strongly persistent. The posterior density of α is concentrated on small values, but with all posterior mass bounded away from zero. This indicates that the variance of the returns is not constant over time. The parameters p and c are intricately connected: A low probability of a high variance disturbance should be linked to a higher variance for this (rare) occurrence. The last panel shows the posterior contours for these two parameters; the negative correlation between p and c is apparent. Note that the lower limit on the range of p is not binding, but that some positive probability is found for values of p near the upper bound of 0.3.

To compare the GARCH-mixture model to the basic GARCH model, in Figure 4 the square roots of the unconditional standard deviation components of the models are plotted. The first three panels are for the mixture model. The first panel displays the (prior and posterior) distributions of σ_{GARCH} for the mixture model, see (12); the second contains the distributions of σ_u, see (14); and the third shows the distributions of the implied unconditional standard deviation of the mixture model, i.e. the square root of $E(\sigma_\epsilon^2) = \omega/(1-\delta-\alpha\sigma_u^2)$. This compares to the last panel with the unconditional standard deviation of the simple GARCH model.

From these graphs we see that both models tend to be stationary with unconditional standard deviations around 0.85, in close correspondence with the data standard deviation (0.83). Note that though non-stationarity was ruled out by the prior we applied, large values for the unconditional standard deviation may still occur. Indeed, about 1% of the calculated standard deviations in the mixture model was larger than 1.5; these have not been taken up into the third panel of Figure 4. The unconditional standard deviation is on average smaller in the mixture model, though its distribution is more spread out. The data seem to be quite informative about the parameters of the mixture distributions, indicating that this element of the model helps to explain the variance in the data series. The mixture parameters, estimated at the posterior mean, indicate that on average 15% of the disturbances exhibit a fourfold

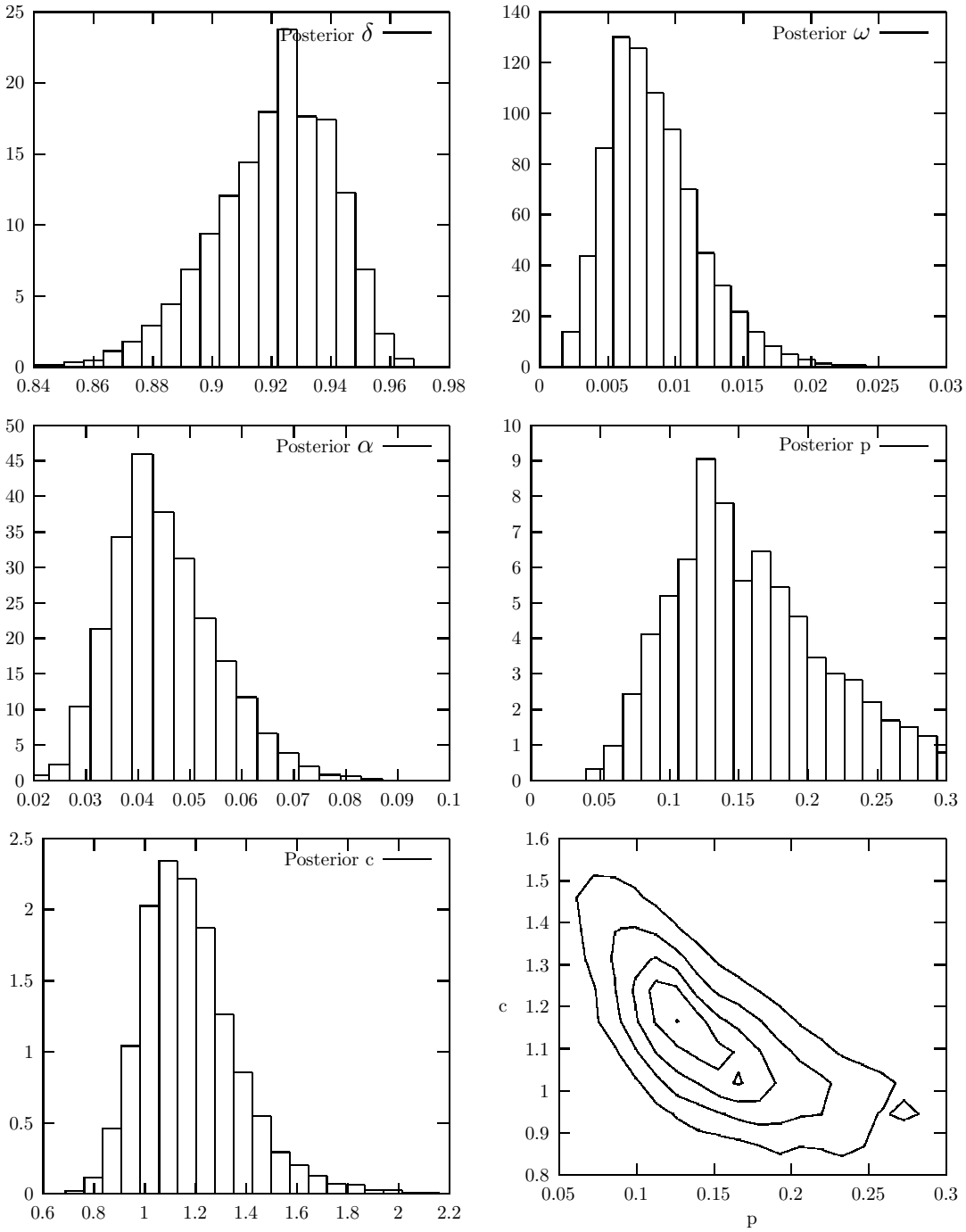

Figure 3: Marginal posterior densities of $\delta, \omega, \alpha, p$, and c, and contours of the joint posterior of p and c

Table 6: Results for the APS algorithm on the Dow Jones data

	GARCH mixture	**GARCH**
δ	0.92 (0.020)	0.89 (0.019)
ω	0.08 (0.003)	0.02 (0.005)
α	0.04 (0.010)	0.09 (0.014)
p	0.15 (0.055)	
c	1.14 (0.189)	
Acc. rate 2	0.70	0.96
Acc. rate 3	0.68	0.97
Acc. rate 4	0.73	0.97
Mahalanobis 2	0.13	0.03
Mahalanobis 3	0.05	0.00
Mahalanobis 4	0.04	0.00
Autocorrelation ρ_1^*	0.37	0.07
Fn. ev. (last round, $\times$1000)	47	31
Fn. ev. (total, $\times$1000)	59	39

In the first block, entries are posterior expectations and standard deviations (between parentheses). The remaining blocks contain the same information as in Table 2.

larger variance. Likewise, the variance of the disturbance term (at the posterior mean) is equal to 1.54, see Equation (14).

In Table 6, the diagnostics explained in Section 3 on the convergence of APS are reported for both models. Not surprisingly, convergence is faster for the simple model. The APS algorithm was started using ML estimates of the parameters. This explains the high acceptance rates in the MH step in the second round.

Predictive analysis and Value-at-Risk

In financial practice the Value-at-Risk (VaR) of a portfolio is often calculated as a measure of risk. The VaR is the maximum loss which can be expected at a fixed confidence level for a fixed horizon, see e.g. Jorion (1997), p 87-88. More precisely, the VaR at horizon τ, with confidence $1-\alpha$, of a $\$W$ initial investment is defined as

$$\text{VaR}(\alpha, \tau) = W[E(s_{t+\tau}) - s_\alpha]$$

where $s_{t+\tau}$ is the return at time $t+\tau$ on an investment of \$1 made at time t, and s_α is the α%-quantile of the distribution of $s_{t+\tau}$, i.e. the solution of

$$\int_{s_\alpha}^{\infty} f(s_{t+\tau}) ds_{t+\tau} = \alpha$$

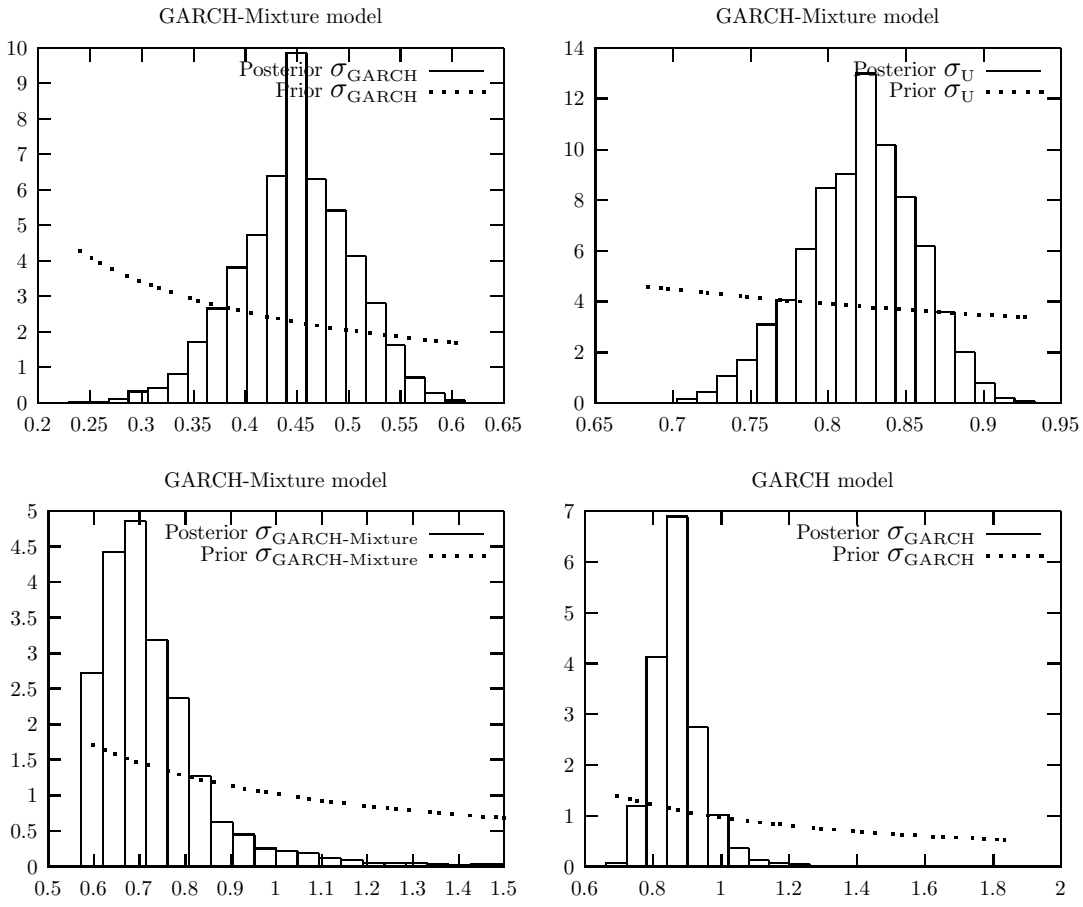

Figure 4: Prior and posterior densities of standard deviations for the GARCH-mixture and the simple GARCH models

with $f(s_{t+\tau})$ the density of the corresponding return.

With a given model, say the GARCH model, the VaR is a function of the model parameters and of future returns. Bayesian inference on the VaR implies that one integrates over the parameters and over the future returns, i.e. to compute VaR from the predictive distribution of the future return $s_{t+\tau}$. It is convenient to make use of sampling methods in this case; see Geweke (1989) for a predictive analysis with an ARCH-model using importance sampling.

This is implemented as follows: we generate (by APS) a parameter value $\theta = (\delta, \omega, \alpha, p, c)$ from the posterior distribution, and we generate from the conditional (on θ) predictive density a random sample of returns for $1, 2, .., \tau$ periods ahead. Repeated N times, this procedure provides for each horizon a random sample of N returns generated from the predictive distribution. For a \$1 investment, the VaR at horizon τ with confidence $1-\alpha$ is then approximated by the average of the generated sample of $s_{t+\tau}$ less the $\alpha\%$-quantile of the same sample.

The simulation which is needed for evaluating the Value-at-Risk is computer-intensive. Therefore it is advisable to start with a sample of posterior parameters θ which is of good quality, showing little serial correlation, such that only a 'small' sample suffices for the calculation of the VaR. APS in such a case is a good choice, as it tends to result in posterior samples with less correlation than its direct competitors.

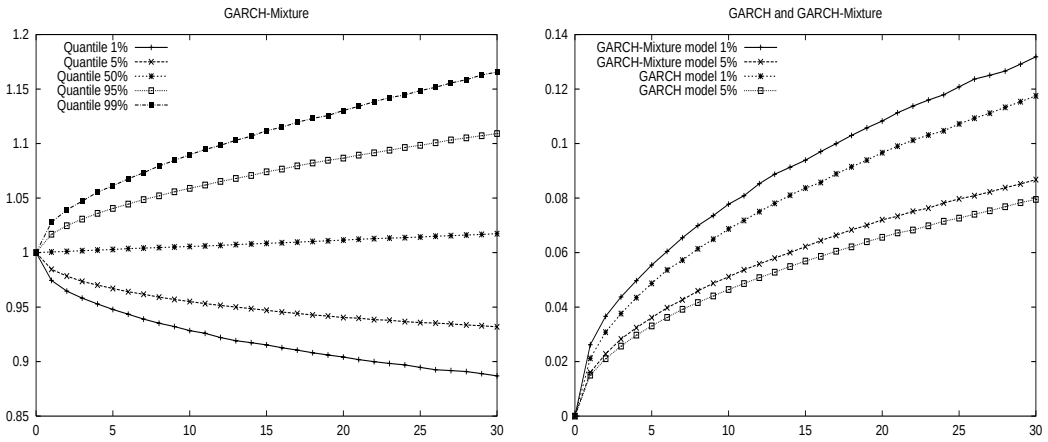

Figure 5: Quantiles of return-on-investment for the GARCH-mixture model over different horizons (left panel) and VaR evaluated for both the GARCH and the GARCH-mixture models (right panel)

This sampling method was executed on both the GARCH and GARCH-mixture models for the returns on the Dow Jones index. Outcomes are shown in Figure 5. The left panel

depicts the evolution of the 1, 5, 50, 95 and 99% quantiles of the return-on-investment in the index over 30 time periods into the future assuming the GARCH-mixture model holds. The right panel shows $VaR(\alpha, \tau)$, calculated as indicated above for the GARCH and the GARCH-mixture models, for $\alpha = 1$ and 5% and over the 30 trading days horizon.

It is seen on Figure 5 (right panel) that the GARCH model underestimates the volatility in the returns on larger horizons, resulting in an estimated 1% VaR at a 30 day horizon given as 0.117 instead of 0.132. The fact that the GARCH-mixture model assigns a 15% probability of finding a disturbance on the return on investment from a larger variance distribution leads to more conservative risk estimates, which indeed is an important piece of information for an investor.

Another way of investigating the difference in implied VaR for the two models is to compare the confidence levels linked to certain values for the VaR, at fixed horizon. Note that the VaR is a function of the confidence level and the horizon, $\text{VaR} = v(\alpha, \tau|M)$. It then follows that

$$\alpha(\text{VaR}, \tau|M) = v^{-1}(\text{VaR}, \tau|M)$$

In Figure 6 we depict the relation between the confidence levels for the GARCH and GARCH-mixture models,

$$r(\text{VaR}, \tau) = \frac{\alpha(\text{VaR}, \tau|\text{GARCH})}{\alpha(\text{VaR}, \tau|\text{GARCH-mixture})}.$$

It is seen that the GARCH model assigns a lower probability to a relatively large possible loss, especially at larger horizons. E.g. for an investment at the largest horizon, the probability of encountering a negative return-on-investment of at least 10% is estimated 25 times higher for the GARCH-mixture than for the GARCH model.

5 Conclusions

In this paper a new MCMC algorithm was introduced: adaptive polar sampling, where sampling does not take place in the n-dimensional parameter space directly, but in an $(n-1)$-dimensional subspace in which the target density is usually better behaved.

In a variety of models which feature ill-behaved posterior distributions, the APS algorithm was found to perform much better, in terms of acceptance rate and autocorrelation structure of the generated samples, than the standard MH algorithm in the n-dimensional space. These

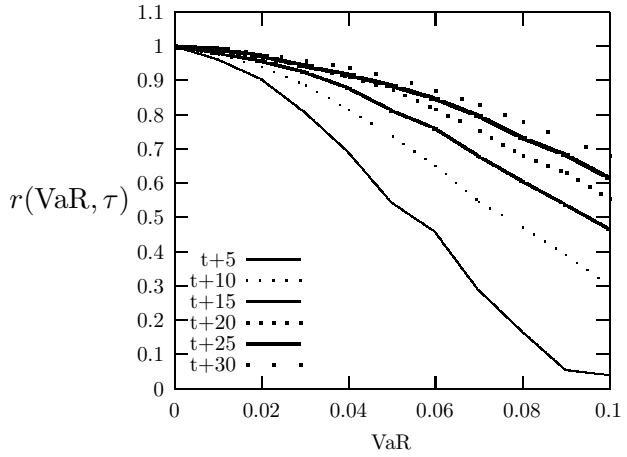

Figure 6: Relation between the implied confidence levels and the Value-at-Risk

advantages of robustness and flexibility of APS come at the cost of a higher computational burden.

The algorithm was applied to sample from the posterior distribution of the parameters of both a simple GARCH model and a more elaborate GARCH-mixture model, on a dataset concerning the daily returns on the Dow Jones stock index. The application was aimed at evaluating the Value-at-Risk implied by the models. It was found that the GARCH-mixture model is favoured by the data compared to a standard GARCH model. For the Value-at-Risk, it was found that the difference between the elaborate and parsimonious model is less than 1% at a time horizon of 30 trading days, when the 5% VaR is concerned. The 1% VaR was calculated with a larger gap than 1% of the invested sum. This difference indeed is important information for an investor.

We end the paper with a remark. The combined use of the polar transformation and the deterministic integration on the distances enables one to construct four Monte Carlo integration algorithms. First, we introduced APS where MH is applied on the directions and distances are generated randomly from its, numerically determined, distribution function. Second, there exists MIXIN (van Dijk et al. 1985) where importance sampling is applied to the direction and Gaussian quadrature to the distances. Third, one may combine importance sampling on the directions with randomly generated distances as in APS. Fourth, a combination of an MH step on directions with deterministic integration on ρ would lead to a viable alternative algorithm. It is a matter of further research to compare these four Monte Carlo

integration schemes.

Further research is also needed to compare APS to some recent sampling algorithms, like the simulated tempering of Marinari and Parisi (1992) and the simulated sintering of Liu and Sabatti (1998), and to experiment with APS in other classes of models, such as reduced rank models, see e.g. Kleibergen and van Dijk (1998).

Appendix

In Section 2 the marginal density of η resulting from a normal density in the original parameter space is used. Here, we present the derivation of the transformed candidate and target densities. Denote the normal density in the original space by

$$q(\theta) \propto \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta-\mu)'\Sigma^{-\frac{1}{2}}(\theta-\mu)\right),$$

and denote by $q(\eta,\rho), q(\eta)$ and $q(\rho)$ the joint and marginal densities of η and ρ after the polar transformation defined by (4)-(5). The following results hold:

$$\begin{aligned} q(\eta,\rho) &= |J(\eta,\rho,\Sigma)|\, q(T^{-1}(\eta,\rho)) \\ &\propto \det(\Sigma)^{\frac{1}{2}} |\rho|^{n-1} |\prod_{i=1}^{n-2} \cos^{n-i-1}\eta_i| \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta-\mu)'\Sigma^{-\frac{1}{2}}(\theta-\mu)\right) \\ &= |\prod_{i=1}^{n-2} \cos^{n-i-1}\eta_i| \times |\rho|^{n-1} \times \exp\left(-\frac{1}{2}\rho^2\right) \\ &= |J(\eta)| \times |J(\rho)| \times \exp\left(-\frac{1}{2}\rho^2\right) \end{aligned} \tag{16}$$

$$q(\eta) \propto |J(\eta)| \tag{17}$$

$$q(\rho) \propto |J(\rho)| \exp\left(-\frac{1}{2}\rho^2\right), \tag{18}$$

see also Equation 7. For the target density, the independence between ρ and η does not hold in general. The marginal density of η for the target density $p_\theta(\theta)$ is

$$p(\eta) = \int p_\theta\left(T^{-1}(\eta,\rho)\right) |J(\eta,\rho,\Sigma)|\, d\rho = |J(\eta)|\, |J(\Sigma)| \int p_\theta\left(T^{-1}(\eta,\rho)\right) |J(\rho)|\, d\rho. \tag{19}$$

Acknowledgement: The estimations and simulations reported in this article have been calculated using programs written by the authors. Programs were written in Gauss (version 3.2.15) as well as in Ox (version 2.10, see Doornik 1998).

References

Albert, J. H. & Chib, S. (1993), 'Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts', *Journal of Business & Economic Statistics* **11**, 1–16.

Bauwens, L. & Lubrano, M. (1998), 'Bayesian inference on GARCH models using the Gibbs sampler', *The Econometrics Journal* pp. C23–C46.

Chib, S. & Greenberg, E. (1995), 'Understanding the Metropolis-Hastings algorithm', *The American Statistician* **49**, 327–335.

Chib, S. & Greenberg, E. (1996), 'Markov Chain Monte Carlo simulation methods in econometrics', *Econometric Theory* **12**, 409–431.

Diebolt, J. & Robert, C. P. (1994), 'Estimation of finite mixture distributions through Bayesian sampling', *Journal of the Royal Statistical Society* **56**, 363–376.

Doornik, J. A. (1998), *Object-Oriented Matrix Programming using Ox 2.0*, London: Timberlake Consultants Ltd. See `http://www.nuff.ox.ac.uk/Users/Doornik`.

Engle, R. F. (1995), *Arch: Selected Readings*, Advanced Texts in Econometrics, Oxford [etc.]: Oxford University Press.

Gelfand, A. E., Hills, S. E., Racine-Poon, A. & Smith, A. F. M. (1990), 'Illustation of Bayesian inference in normal data models using Gibbs sampling', *Journal of the American Statistical Association* **85**, 972–985.

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

Geweke, J. (1992), Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J.-M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds, 'Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting', Oxford: Clarendon Press, pp. 169–193.

Geweke, J. (1993), 'Bayesian treatment of the independent Student-t linear model', *Journal of Applied Econometrics* **8**, S19–S40.

Geweke, J. (1999), Using simulation methods for Bayesian econometric models: Inference, development, and communication. Forthcoming in Econometric Reviews.

Givens, G. H. & Raftery, A. E. (1996), 'Local adaptive importance sampling for multivariate densities with strong nonlinear relationships', *Journal of the American Statistical Association* **91**, 132–141.

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

Hobert, J. P. & Casella, G. (1996), 'The effect of improper priors on Gibbs sampling in hierarchical linear mixed models', *Journal of the American Statistical Association* **91**, 1461–1473.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models', *Journal of Business and Economic Statistics* **12**, 371–417.

Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, New York: McGraw-Hill.

Justel, A. & Peña, D. (1996), 'Gibbs sampling will fail in outlier problems with strong masking', *Journal of Computational and Graphical Statistics* **5**, 176–189.

Kendall, M. G. & Stuart, A. (1973), *The Advanced Theory of Statistics*, Vol. 1: Distribution Theory, 3rd edition edn, London: Griffin.

Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: Likelihood inference and comparison with ARCH models', *Review of Economic Studies* **64**, 361–393.

Kleibergen, F. R. & van Dijk, H. K. (1993), 'Non-stationarity in GARCH models: A Bayesian analysis', *Journal of Applied Econometrics* **8**, S41–S61.

Kleibergen, F. R. & van Dijk, H. K. (1994), 'On the shape of the likelihood/posterior in cointegration models', *Econometric Theory* **10**, 514–551.

Kleibergen, F. R. & van Dijk, H. K. (1998), 'Bayesian simultaneous equations analysis using reduced rank structures', *Econometric Theory* **14**, 701–743.

Koop, G. & van Dijk, H. K. (1999), 'Testing for integration using evolving trend and seasonal models: A Bayesian approach', *Forthcoming in Journal of Econometrics* .

Liu, J. S. & Sabatti, C. (1998), 'Simulated sintering: Markov Chain Monte Carlo with spaces of varying dimensions', *Proceedings of the 6th International Meeting on Bayesian Statistics, Valencia, Spain* .

Marinari, E. & Parisi, G. (1992), 'Simulated tempering: A new Monte Carlo scheme', *Europhysics Letters* **19**, 451–458.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equations of state calculations by fast computing machines', *Journal of Chemical Physics* **21**, 1087–1091.

Nelson, D. B. (1990), 'Stationarity and persistence in the GARCH(1, 1) model', *Econometric Theory* **6**, 318–334.

Oh, M. S. & Berger, J. O. (1992), 'Adaptive importance sampling in Monte Carlo integration', *Journal of Statistical Computation and Simulation* **41**, 143–168.

Paap, R. & van Dijk, H. K. (1999), Bayes estimates of Markov trends in possibly cointegrated series: An application to US consumption and income, Technical Report EI-9911/A, Econometric Institute, Erasmus University Rotterdam.

Roberts, G. O. & Smith, A. F. M. (1994), 'Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms', *Stochastic Processes and their Applications* **49**.

Rubinstein, R. (1981), *Simulation and the Monte Carlo Method*, New York: Wiley.

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

van Dijk, H. K. & Kloek, T. (1980), 'Further experience in Bayesian analysis using Monte Carlo integration', *Journal of Econometrics* **14**, 307–328.

van Dijk, H. K., Kloek, T. & Boender, C. G. E. (1985), 'Posterior moments computed by mixed integration', *Journal of Econometrics* **29**, 3–18.