ADAPTIVE POLAR SAMPLING:
A class of flexible and robust Monte Carlo integration methods

Luc Bauwens[1], Charles S. Bos[2], Herman K. van Dijk[3], and Rutger D. van Oest[3]

Econometric Institute Report EI 2002-27

Abstract

Adaptive Polar Sampling (APS) algorithms are proposed for Bayesian analysis of models with nonelliptical, possibly, multimodal posterior distributions. A location-scale transformation and a transformation to polar coordinates are used. After the transformation to polar coordinates, a Metropolis-Hastings method or, alternatively, an importance sampling method is applied to sample directions and, conditionally on these, distances are generated by inverting the cumulative distribution function. A sequential procedure is applied to update the initial location and scaling matrix in order to sample directions in an efficient way.

Tested on a set of canonical mixture models that feature multimodality, strong correlation, and skewness, the APS algorithms compare favourably with the standard Metropolis-Hastings and importance samplers in terms of flexibility and robustness. APS is applied to several econometric and statistical examples. The empirical results for a regression model with scale contamination, an ARMA-GARCH-Student t model with near cancellation of roots and heavy tails, a mixture model for economic growth, and a nonlinear threshold model for industrial production growth confirm the practical flexibility and robustness of APS.

Keywords: Markov chain Monte Carlo, importance sampling, polar coordinates

JEL classification: C11, C15, C63

---

[1]CORE and Department of Economics, Université catholique de Louvain.
[2]Econometrics & O.R. Department, Free University of Amsterdam.
[3]Econometric Institute, Erasmus University Rotterdam.

1 Introduction

In recent decades Markov Chain Monte Carlo (MCMC) methods, in particular Metropolis-Hastings (MH) and Gibbs sampling (GS) and, to a lesser extent, indirect sampling methods like importance sampling (IS), have been applied extensively and successfully 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). Importance sampling, see Hammersley and Handscomb (1964), was introduced in Bayesian analysis by Kloek and Van Dijk (1978) and further developed by Van Dijk and Kloek (1980,1984), and by Geweke (1989).

Although MC methods revolutionized the applicability of Bayesian inference, there is, in practice, a substantial variation in their convergence behaviour. The special features of the sampling method, the complex structure of the model, or the nature of the data may be the culprit of such behaviour. Hobert and Casella (1996) show for instance that the Gibbs sampler does not converge for the case of a hierarchical linear mixed model when the prior is uniform. Other examples of complex models are the ones with reduced rank structures. Kleibergen and Van Dijk (1994,1998) demonstrate near reducibility of MCMC methods when there exists near nonidentifiability and nonstationarity in econometric models with flat priors. Justel and Peña (1996) emphasize the convergence problems of the Gibbs sampler when there are outliers in the data. The performance of the Gibbs sampler is also seriously hampered by strong correlation in the target distribution. Convergence problems of importance sampling using a simple normal or Student t candidate density have been documented by Van Dijk and Kloek (1984) and Geweke (1989). A multimodal target density may pose problems to all methods. If the MH candidate density is unimodal, with low probability of drawing candidate values in one of the 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. Using a unimodal normal or Student t candidate function the method of importance sampling ends up with many drawings having only negligible weights.

A common difficulty encountered in all samplers is the choice of a candidate or importance 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]

In this paper we introduce the class of adaptive polar sampling (APS) methods to sample from a target (posterior) distribution which is possibly multi-modal, skew, and exhibits strong correlation, in summary it is nonelliptical. The APS algorithms feature two transformations to induce a more regular shape of the target function in the transformed space than in the original space. The key transformation is one where the m-dimensional space is transformed into polar coordinates which consist of a distance measure and a $(m-1)$-dimensional vector of directions (or angles). A MH or an IS algorithm is applied to sample the directions. Next, the distance is sampled conditionally on the directions, from the (transformed) target density, 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

---

[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).

second order moments obtained in successive rounds of the algorithm. The adaptive procedure is intended to improve the acceptance rate within the MH step and to give a more uniform distribution of the weights in the IS step.

The advantages of the APS algorithms are twofold. Firstly, the algorithms are *flexible* and *parsimonious* in their use of information on the shape of the target density. Only location and scale need to be specified as initial values. Secondly, the algorithm is *robust*: they can handle a large variety of features of target distributions, in particular multimodality, strong correlation, extreme skewness, and heavy tails. We claim that the APS algorithms avoid the often time-consuming and difficult task, especially for non-experts, of choosing and tuning a sampling algorithm for a specific application, such as computation of posterior moments in Bayesian inference. They can be applied without the need to study in depth the shape of the posterior density in order to design a sophisticated approximation to it. It can be argued that for any specific model and data combination, a more efficient algorithm than APS may be designed, but our viewpoint is that the extra effort (in research time) required to achieve this may not be rewarding.

The APS algorithms extend earlier methods, that is, the mixed integration method by Van Dijk, Kloek, and Boender (1985) and the spherical integration method by Monahan and Genz (1997). In the APS framework one generates a set of (weighted) posterior drawings and one can easily compute any function of the parameters. This in contrast to the earlier methods where univariate integrals are generated, which are cumbersome to handle in further posterior or predictive analysis. The APS methods also generalize Adaptive Direction Sampling (ADS) algorithms proposed by Gilks, Roberts, and George (1994). ADS methods are, in a certain sense, an extension of Gibbs sampling. In ADS, directions are sampled in the original parameter space using only information on the shape of the target. In APS, a MH or IS step is used comparing target and candidate densities in a transformed subspace where the transformed target density is supposed to be more regular than in the original space. Further, in APS, the distances are generated from the exact target.

The outline of the paper is as follows. In Section 2 the algorithms are introduced. In Section 3 canonical mixture models are used for experimenting with APS and for comparing its performance with that of the standard Metropolis-Hastings, importance sampling and Gibbs algorithms. The models include mixtures with multimodality, extreme correlation and skewness. We also apply APS to several empirical examples to illustrate its practical usefulness. In Section 4 we use a regression model with scale contamination in order to investigate a study from Justel and Peña (1996) concerning the oxidation of ammonia to nitric acid in a plant. Next, we investigate a study from Bos, Mahieu, and Van Dijk (2000) on predictability of daily exchange rate data using an ARMA-GARCH Student t model which has near cancellation of roots and heavy tails. Third, we analyze economic growth of the USA using a mixture model. Finally, we make use of a threshold model to study the nonlinear time series features of the industrial production in the USA. Conclusions are presented in Section 5 and technical details are given in the appendices.

2   Adaptive polar sampling

Most simulation algorithms for posterior distributions generate random drawings in the original parameter space. Several researchers advocate to simulate in a transformed space, where the simulation is more efficient in some sense, see e.g. Gilks and Roberts (1996). For example,

if there exists a strong correlation between two random variables, an orthogonalising transformation reduces serial dependence in a Gibbs sampling scheme. Another example arises in the context of importance sampling: when an adequate transformation yields a distribution that is much closer to a symmetric one than the original one, an efficient importance function is easier to construct. The adaptive polar sampling algorithms rely on this general idea. They are based on a transformation to polar coordinates. Heuristically, the original parameter space is transformed into a $(m-1)$-dimensional space of angles or directions in which the density is assumed to be more well behaved, and a unidimensional complementary space in which most of the variation (or ill behavior) of the target density is concentrated. We note that when the target or candidate density is a member of the elliptical family one can make use of the following result. The density of the $(m-1)$-dimensional directions or angles is uniform on the unit sphere and this density is independent of the density of the unidimensional distance which has a known analytical form (e.g. a member of the gamma family when the candidate is normal); see e.g. Muirhead (1982) (section 1.5). This result is the basis for the Box-Muller method of generating normal random variables; see below (3) and (4) and Box and Muller (1958).

In this section, we concentrate on two members of the polar sampling algorithms, one is of the Metropolis-Hastings type, and the second one is of the importance sampling type. These variants will be referred to as adaptive polar Metropolis-Hastings (APMH) and adaptive polar importance sampling (APIS), respectively. Since the signed polar transformation is at the heart of the algorithms, we start with a description of this transformation and its relation to the well-known standard polar transformation.

2.1 Signed polar transformation

2.1.1 2-dimensional signed polar transformation

In usual polar coordinates, the position of a point in R^2 is described by the Euclidian distance from the origin to the point and by the angle formed by the horizontal axis and the line from the point to the origin. The distance is positive and the angle takes a value between 0 and 2π (spanning a full circle). The standard polar transformation can be found in Muirhead (1982). The signed polar transformation is a convenient modification of the polar transformation: the distance is positive if the point is on the right of the vertical axis, negative otherwise, and the angle takes a value between $[-\pi/2, \pi/2]$ (spanning a half circle). The situation is illustrated on Figure 1. For a point $y = (y_1, y_2)' \in R^2$, the signed polar transformation is defined by

$$\rho(y_1, y_2) \;=\; \mathrm{sgn}(y_1)\sqrt{y_1^2 + y_2^2} \;\in R, \tag{1}$$

$$\eta(y_1, y_2) \;=\; \arcsin\!\left(\frac{y_2}{\rho(y_1, y_2)}\right) \;\in [-\frac{\pi}{2}, \frac{\pi}{2}], \tag{2}$$

where $\mathrm{sgn}(y_1) = 1$ if $y_1 \geq 0$ and $= -1$ if $y_1 < 0$.

The inverse transformation is given by

$$y_1(\rho, \eta) \;=\; \rho\cos(\eta), \tag{3}$$

$$y_2(\rho, \eta) \;=\; \rho\sin(\eta), \tag{4}$$

and the Jacobian is given by

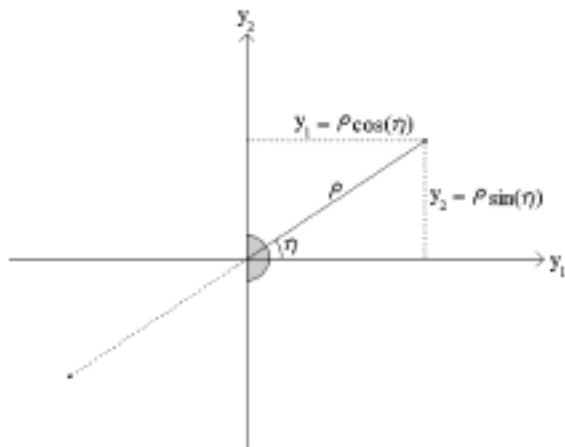$$J_{y_1, y_2}(\rho, \eta) = \rho. \tag{5}$$

3

Figure 1: Signed polar coordinates.

2.1.2 m-dimensional signed polar transformation

In the m-dimensional case, the distance and $m-1$ angles describe the position of a point in R^m. The m-dimensional signed polar transformation from $y = (y_1, y_2, \ldots, y_m)' \in R^m$ to $(\rho, \eta_1, \ldots, \eta_{m-1})' \in R \times [-\frac{\pi}{2}, \frac{\pi}{2}]^{m-1}$ is defined by

$$\rho(y) = \text{sgn}(y_1)\sqrt{y'y}, \tag{6}$$

$$\eta_j(y) = \arcsin\left(\frac{y_{m-j+1}}{\rho(y)\prod_{i=1}^{j-1}\cos(\eta_i(y))}\right), \qquad j = 1, 2, \ldots, m-1, \tag{7}$$

where by convention, $\prod_{i=1}^{0}\cos(\eta_i) = 1$. It can be derived from (6) and (7) that the transformation, the other way around, is defined by

$$y_1(\rho, \eta) = \rho \prod_{i=1}^{m-1}\cos(\eta_i), \tag{8}$$

$$y_j(\rho, \eta) = \rho \sin(\eta_{m-j+1}) \prod_{i=1}^{m-j}\cos(\eta_i), \qquad j = 2, \ldots, m. \tag{9}$$

The Jacobian of the transformation is given by

$$J_y(\rho, \eta) = \rho^{m-1} \prod_{i=1}^{m-2} \cos^{m-i-1}(\eta_i) \equiv J_y(\rho)J_y(\eta). \tag{10}$$

We end this section with a remark. For expository purposes we make use of the polar transformation. We note that one may also use a transformation using Stiefel manifolds. For details we refer to Muirhead (1982) (see Chapters 1 and 2). The important point is that both transformations induce independence between the directions and the distance in the class of spherical distributions (see Theorems 1.5.5 and 1.5.6 in Muirhead), of which the normal is a special case. The geometrical interpretation of polar coordinates may be easier to understand than the more abstract analysis using Stiefel manifolds.

4

2.2 Adaptive polar Metropolis-Hastings

We start by defining the polar Metropolis-Hastings algorithm (PMH), which is based on a candidate generating density that is taken to be multivariate normal with parameters μ and Σ. Next, we illustrate the algorithm for a bimodal target density. Finally, we define the adaptive PMH algorithm (APMH), where μ and Σ are updated using the sample of draws from a previous round of the PMH algorithm.

2.2.1 Definition of PMH

PMH is based on an independence chain MH algorithm. It uses draws from a $N(\mu, \Sigma)$ candidate where hopefully μ and Σ provide good approximations to the unknown mean and covariance matrix of the target distribution (see Subsection 2.2.3). We note that normality of the candidate density is only relevant to the extent that drawings should be generated from a member of the class of elliptical distributions; see also remark 1 below. In contrast with the MH algorithm, the drawings are not used for construction of a Markov chain in the original parameter space. Instead, two transformations are made.

The first transformation is a location-scale transformation, aiming at standardizing the candidate density with respect to the location, scale, and correlations of the target (posterior) density, denoted by $p(x)$. The location-scale transformation is given by[2]

$$y = y(x \,|\, \mu, \Sigma) = \Sigma^{-1/2}(x - \mu),\tag{11}$$

with inverse transformation

$$x = x(y \,|\, \mu, \Sigma) = \mu + \Sigma^{1/2} y,\tag{12}$$

and Jacobian

$$J_x(y) = \det(\Sigma^{1/2}).\tag{13}$$

The second transformation is the signed polar transformation, which is defined by (6) and (7), with Jacobian (10).

Combining the two transformations, one obtains the composite transformation

$$\begin{pmatrix} \rho \\ \eta \end{pmatrix} = \begin{pmatrix} \rho(x \,|\, \mu, \Sigma) \\ \eta(x \,|\, \mu, \Sigma) \end{pmatrix} = \begin{pmatrix} \rho(y(x \,|\, \mu, \Sigma)) \\ \eta(y(x \,|\, \mu, \Sigma)) \end{pmatrix}\tag{14}$$

with inverse transformation

$$x = x(\rho, \eta \,|\, \mu, \Sigma) = x(y(\rho, \eta) \,|\, \mu, \Sigma)\tag{15}$$

and Jacobian

$$J_x(\rho, \eta) = J_y(\rho, \eta) J_x(y) = J_y(\rho) J_y(\eta) \det(\Sigma^{1/2}).\tag{16}$$

Applying the two transformations to a candidate realization x_i from $\sim N(\mu, \Sigma)$ yields a distance ρ_i^* and a vector of angles η_i^* (referred to hereafter as a 'direction').[3] Ignoring the

---

[2]$\Sigma^{1/2}$ denotes the Cholesky decomposition of Σ, and $\Sigma^{-1/2}$ denotes the inverse matrix of $\Sigma^{1/2}$.

[3]From here on, the index i (in x_i^*, η_i^* ...) does not indicate the i-th element of the corresponding vector, but indicates the number of the draw in a sequence of successive draws.

distance, the candidate direction is either accepted or rejected in a MH step, that is, the direction becomes

$$\eta_i = \begin{cases} \eta_i^* & \text{with probability } \alpha(\eta_{i-1}, \eta_i^*) \\ \eta_{i-1} & \text{with probability } 1 - \alpha(\eta_{i-1}, \eta_i^*) \end{cases} \tag{17}$$

for some acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, which is given in Proposition 1 below. An iteration of APMH is completed by drawing from the target distribution on the line defined by the direction η_i. This can be done as follows. First, one draws a distance ρ_i from the transformed target distribution for given direction η_i using the numerical inverse transform method, see Proposition 1. Next, ρ_i and η_i are transformed to the original space by inverting the signed polar transformation and the location-scale transformation. In Table 1, we summarize the steps of one iteration of PMH.

Table 1: One iteration of PMH

---
1. Generate x_i^* from $N(\mu, \Sigma)$
2. Transform x_i^* to $y_i^* = \Sigma^{-1/2}(x_i^* - \mu)$
3. Transform y_i^* to ρ_i^* and η_i^*, using (6) and (7)
4. Apply MH step to η_i, see (17)
5. Generate ρ_i from $p(\rho|\eta_i)$ by inverting numerically its cdf
6. Transform ρ_i and η_i to y_i, using (8) and (9)
7. Transform y_i to $x_i = \mu + \Sigma^{1/2} y_i$

Note that steps 1 and 2 amount to generating y_i^* from $N(0, I_m)$. We want to make explicit the dependence on μ and Σ.

Step 4 of a PMH iteration requires the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, and step 5 requires the distribution of the distance ρ conditional on the direction η_i. They are given in the next proposition.

Proposition 1 *The acceptance probability of step 3 of the PMH algorithm summarized in Table 1 is given by*

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{ \frac{I(\eta_i^*)}{I(\eta_{i-1})}, 1 \right\}, \tag{18}$$

where

$$I(\eta) = \int_{-\infty}^{\infty} \kappa(\rho|\eta) \, d\rho, \tag{19}$$

where $\kappa(\rho|\eta)$ is a kernel of the the conditional density $p(\rho|\eta)$ of step 4, which is defined by

$$p(\rho|\eta) \propto \kappa(\rho|\eta) = p(x(\rho, \eta| \mu, \Sigma)) \, |J_y(\rho)|. \tag{20}$$

Proof: See Appendix 1. ⋄

Remark 1: A noteworthy property is that the acceptance probability does not depend on the functional form of the candidate density under the condition that this candidate density

6

is of the form $f\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)$, i.e. an elliptically-contoured density. However, the acceptance probability depends on the generated direction η and thus on the location and scaling matrix of the candidate density.

Remark 2: In order to obtain the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, the integral $I(\eta)$ defined by (19) can be computed by a deterministic integration rule. Since the density of ρ conditional on η is proportional to the integrand of $I(\eta)$, evaluations of the integrand, gathered during the deterministic integration phase, can be used in order to construct a grid for $p(\rho|\eta)$. Using the numerical inverse transform method, sampling the distance ρ conditional on the direction η, that is, step 5 of a PMH iteration, is straightforward. One may interpret step 5 as a "Griddy Gibbs" step

Remark 3: Theoretical convergence of PMH is quite intuitive. The usual sufficient condition for the MH algorithm may be adapted to the condition that $q(\eta) > 0$ for all η such that $p(\eta) > 0$. Since $q(\eta) \propto J_y(\eta) = \prod_{i=1}^{m-2} \cos^{m-i-1}(\eta_i) > 0$ for all $\eta \in [-\frac{\pi}{2}, \frac{\pi}{2}]^{m-1}$, this condition is satisfied.

In practice we reduce the computational effort by generating several drawings of ρ for each drawing of η, i.e. we capitalize on the construction of $p(\rho|\eta)$ (see remark 2). Note that the computed integrals (MC estimators) still converge to the theoretical integrals. The main point is that although the generated drawings of y and x are dependent, the computed integrals are consistent estimates of the theoretical values of the integrals that one is interested in, see Geweke (1999) (p 44) and the references cited there.

2.2.2 Illustration

Figure 2 illustrates PMH for a bivariate bimodal target distribution. The upper two graphs display the target density in the original space. A point, representing a realization from the normal candidate distribution $N(\mu, \Sigma)$, is visible in the contour plot. If μ and Σ coincide with the mean and the covariance matrix of the target distribution, then the location-scale transformation leads to the target density that is depicted in the middle graphs. The gain of the location-scale transformation is clear: the density mass is better located around the origin in the sense that a line through the origin, defined by some direction η, 'hits the density mass' more easily. Since PMH precisely considers such lines, the location-scale transformation may lead to a substantial improvement for appropriate μ and Σ. The target density after applying the signed polar transformation is depicted in the bottom two graphs. Although the transformed target density is ill-behaved with respect to ρ, it is well-behaved with respect to η.

Seven steps are distinguished in an iteration of APMH. In step 1, the point in the upper contour plot is obtained from $N(\mu, \Sigma)$ and this point is transformed in step 2 to the point in the middle contour plot. Step 3 results in the point in the bottom contour plot. Now, assume that the direction η is accepted in step 4, then step 5 consists of drawing a (or several) distance(s) ρ along the vertical line. Step 6 can be represented by the transformation of points generated on the line in the bottom contour plot to points generated on the line in the middle contour plot. Similarly, step 7 results in points generated on the line in the upper contour plot.
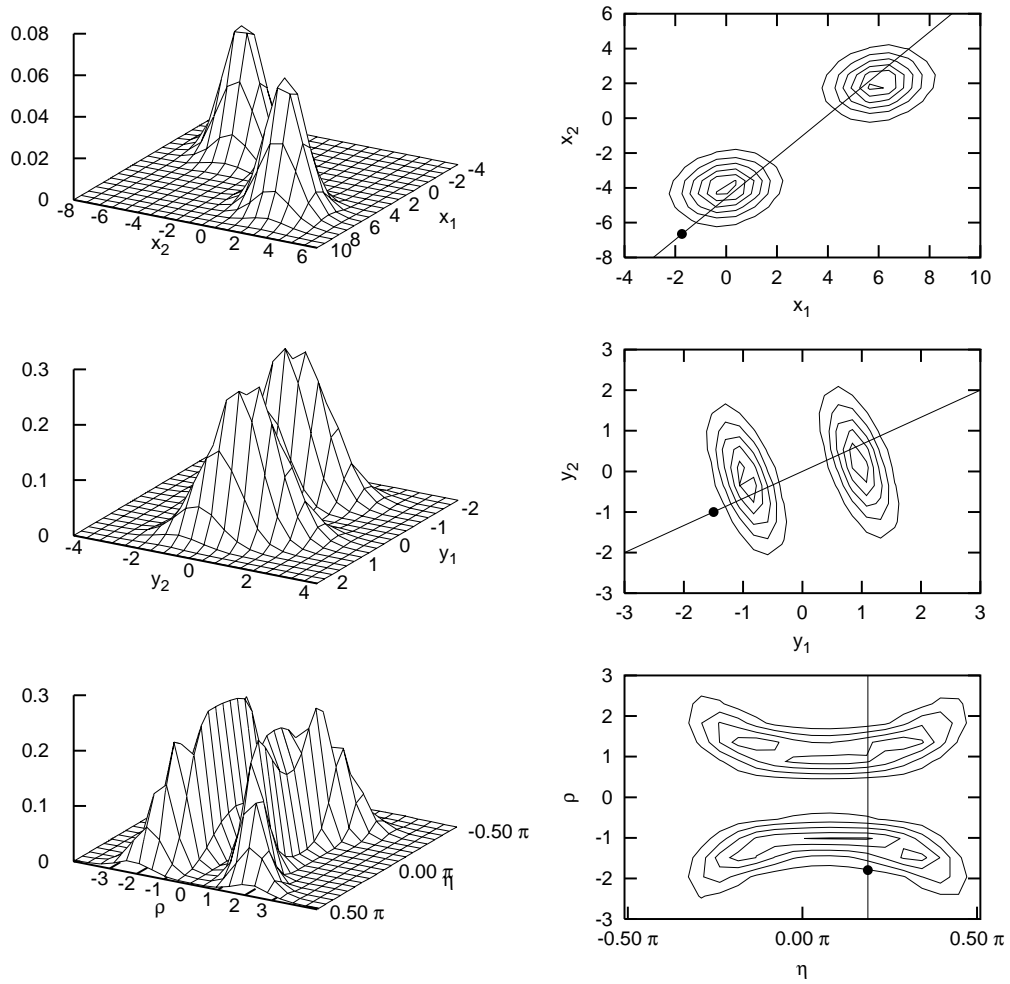
Figure 2: Adaptive polar sampling: target density in original space (above), target density after location-scale transformation (middle) and target density after signed polar transformation (below).

2.2.3 Adaptive polar Metropolis-Hastings

The mean μ and the covariance matrix Σ of the normal candidate distribution have to be specified. Good enough initial approximations are usually the posterior mode and minus the inverse Hessian of the log posterior evaluated at the mode. Heuristically, convergence of PMH should improve if μ and Σ are close to, rather than far from, the target mean and covariance matrix, respectively. APMH considers a sequential adaptive approach. Given a generated sample $x_1, x_2, \ldots, x_n$ from a previous run of the algorithm, μ and Σ are replaced by the Monte Carlo estimates of $E(x)$ and $Cov(x)$, which are given by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i, \tag{21}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})', \tag{22}$$

respectively. Using these estimates, one can proceed with a new sampling round. This process can be repeated any number of times. The initial values of μ and Σ should become less relevant, as they are updated anyway. A danger of the adaptive approach is that information, coming from a 'wrong' sample, may have a misleading effect and may worsen convergence. However, this is not very likely and convergence should be monitored by usual tools. Moreover, since only the direction η, and not the distance ρ, depends on the candidate distribution, the risk of collecting a 'wrong' sample is reduced. PMH should be quite robust, as the distance ρ conditional on the direction η immediately comes from the target distribution, that is, sampling on a given line mimics exactly the target density.

2.3 Adaptive polar importance sampling

Polar importance sampling (PIS) replaces the MH step of PMH for the direction η by an importance sampling step. Every sampled direction is kept, a signed distance is sampled conditional on it, and the resulting polar coordinates are transformed to a draw x in the original space, which is weighted according to the importance weight. A draw x in the original space is a function of a draw (ρ, η) in the transformed space, see (15), implying that the importance weight of (ρ, η) is also the importance weight of x. The importance function of (ρ, η) is defined as

$$q_{imp}(\rho, \eta) = q(\eta)p(\rho|\eta), \tag{23}$$

where $q(\eta)$ and $p(\rho|\eta)$ are defined by (38) and (20) respectively. The corresponding importance weight $w(\rho, \eta)$ is given by

$$w(\rho, \eta) = \frac{p(\rho, \eta)}{q_{imp}(\rho, \eta)} = \frac{p(\eta)p(\rho|\eta)}{q(\eta)p(\rho|\eta)} = \frac{p(\eta)}{q(\eta)} \propto I(\eta) = w(\eta), \tag{24}$$

where $I(\eta)$ is given by (19). As in PMH, the ratio of the target to the candidate density does not depend on ρ since an exact candidate is used for ρ given η.

An interpretation of PIS is that one samples from the target distribution on lines with directions being derived from the candidate distribution. Each line receives a weight, indicating the importance of the underlying direction. The weight of a line is carried over to any realization on that line.

Similar to APMH, the parameters μ and Σ of the location-scale transformation can be updated by replacing them by their Monte Carlo estimates. These estimates are given by

$$\hat{\mu}_w = \frac{\sum_{i=1}^n w(\eta_i) x_i}{\sum_{i=1}^n w(\eta_i)}, \tag{25}$$

$$\hat{\Sigma}_w = \frac{\sum_{i=1}^n w(\eta_i)(x_i - \hat{\mu}_w)(x_i - \hat{\mu}_w)'}{\sum_{i=1}^n w(\eta_i)}, \tag{26}$$

where $x_1, x_2, \ldots, x_n$ is the collected sample, and $w(\eta_1), w(\eta_2), \ldots, w(\eta_n)$ are the corresponding importance weights.

2.4 On related methods

The standard polar transformation is the basis of the well known Box and Muller (1958) method for generating normal random variables. Consider equations (3) and (4), a direction η is generated from a uniform distribution and a distance ρ is generated through a simple inverse transformation from a uniform distribution, see e.g. Rubinstein (1981) (p. 86-87). The APS methods extend the Box-Muller algorithm by generating η using an MH or IS step, where the uniform candidate density is compared with a nonuniform target density. Given a generated candidate η, distances ρ are generated from a very accurate numerical approximation to the target distribution of the distances. Note that this exact distribution is model specific. If the normal candidate density in the original space is a good approximation to the target density in that space then the probability of acceptance in the MH step is close to one and the weight in the IS step is relatively constant. Non-normality can be evaluated using the weights computed in the one-dimensional integration step; see Hop and Van Dijk (1992) and Monahan and Genz (1997).

We emphasize again that within APS one can make use of any candidate that belongs to the family of elliptical distributions. The advantage of the normal is its simplicity and parsimony of parameters: location and scale determine the distribution. A good estimate of the location and scale is important for efficient generation of directions, that is, directions that generate lines which cover the region where the target has substantial probability mass, see e g. the line in Figure 2. We note that Monahan and Genz (1997) use the terminology radial based integration in this context.

The APS class comprises several algorithms. Consider the case where one can distinguish between rejection sampling, importance sampling and Metropolis-Hastings sampling. So far, we have experimented with APMH and APIS. However, one may also define a polar rejection sampling algorithm (APRE): the sampled polar coordinate is accepted if $w(\eta) > cu$ (and rejected otherwise), where u is uniformly drawn in $(0, 1)$, and c is a constant such that the importance function envelopes the target function. For such an algorithm, formulas (21) and (22) apply to the accepted transformed draws. Consider further the case where generating random drawings of ρ is replaced by only evaluating the unidimensional integral. We name this case deterministic integration with respect to ρ. One can combine this deterministic integration with respect to ρ with rejection sampling, importance sampling or Metropolis-Hastings sampling with respect to η and evaluate posterior moments and densities. For the case of importance sampling this has been done in the so-called mixed integration methods of Van Dijk, Kloek, and Boender (1985) and the similar spherical radial integration method of Monahan and Genz (1997). Thus these methods are special cases of the APS class where the step of generating random drawings of ρ is reduced to evaluating only aunidimensional

integral. The limitation of deterministic integration with respect to ρ is that one has to compute a different unidimensional integral for each moment of the target distribution, see Hop and Van Dijk (1992).

Apart from APMH ad APIS we have listed four other members of the APS class. It is also of interest to compare APS with the class of adaptive direction sampling (ADS) algorithms, see Gilks, Roberts, and George (1994). Two well known members of ADS are the hit and run algorithm of Schmaiser and Chen (1991) and the snooker algorithm of Gilks, Roberts, and George (1994). In ADS, directions are sampled in the original parameter space. Only information on the shape of the target density is used. In APS, use is made of an MH or IS step where candidate and target are compared. Further, in APS one generates a distance ρ from a numerically very accurate approximation to the target distribution. This step is not always spelled out ADS. We emphasize that APMH and APIS are members of the MH and IS class of Monte Carlo methods. Convergence properties of these methods are well established. This is not so transparant for the ADS methods.

3   Controlled comparisons of sampling methods

In this section, we compare the performances of APMH and APIS with those of the Metropolis-Hastings algorithm (MH) and importance sampling (IS). We consider two experiments on two 12-dimensional distributions, which are given by

- $\theta \sim N(\mu, \Sigma), \qquad \mu = (1, 2, \ldots, 12)', \qquad \Sigma = 24\, \iota_{12} \iota'_{12} + I_{12},$

- $\theta \sim p_1\, N(\mu_1, \Sigma_1) + p_2\, N(\mu_2, \Sigma_2) + p_3\, N(\mu_3, \Sigma_3), \qquad p_1 = p_2 = p_3 = 1/3,$

  $\mu_1 = -12\, \iota_{12}, \qquad \mu_2 = \begin{pmatrix} -12\, \iota_6 \\ 8\, \iota_6 \end{pmatrix}, \qquad \mu_3 = 8\, \iota_{12}, \qquad \Sigma_1 = \Sigma_2 = \Sigma_3 = 16 I_{12},$

where I_m is the $m \times m$ identity matrix and ι_m is an m-vector consisting of ones. Note that in this section we use the symbol θ instead of x. The first distribution involves quite high correlation, since the correlation is 0.96 for all pairs of components. Features of the second distribution are multimodality (3 modes), skewness, and again high correlation. For comparability, MH and IS are based on a normal candidate distribution and they are made adaptive.

In the experiments, 8 sampling rounds are considered. In each round, APMH and APIS collect 8000 directions and 10 distances in each direction, resulting in a final sample of size 80000. MH and IS are allowed to collect a larger sample of size 800000 in order to make the running times of the four algorithms comparable.[4] The initial mean and the initial covariance matrix of the normal candidate distribution are set at

$$\mu_{\text{init}} = \begin{pmatrix} 4\, \iota_6 \\ -4\, \iota_6 \end{pmatrix}, \qquad \Sigma_{\text{init}} = 200 I_{12}$$

in each experiment. The scale of the covariance matrix is chosen quite large and the correlations are all set to 0. The location of the candidate distribution is far from perfect. In APMH

---

[4]Using Ox (see Doornik 1999), running 8 sampling rounds consumes $20-25$ minutes of computing time on a 550 Mhz Pentium III.

Table 2: Results concerning the high-correlation distribution

	APMH	APIS	MH	IS	true
mean(θ_1)	0.91	0.86	-0.64	-4.13	1.00
mean(θ_2)	1.91	1.88	-1.55	-4.97	2.00
mean(θ_7)	6.91	6.88	4.15	1.09	7.00
mean(θ_8)	7.91	7.87	5.18	4.83	8.00
stdev(θ_1)	4.98	4.96	0.86	0.30	5.00
stdev(θ_2)	4.99	4.95	0.85	0.22	5.00
stdev(θ_7)	4.97	4.95	0.72	0.72	5.00
stdev(θ_8)	4.98	4.95	0.72	0.23	5.00
corr(θ_1,θ_2)	0.96	0.96	0.69	0.67	0.96
corr(θ_6,θ_7)	0.96	0.96	0.49	0.90	0.96
corr(θ_7,θ_8)	0.96	0.96	0.34	0.83	0.96
Mahalanobis	0.00	0.04	1.77	12.69	

and APIS, the integral $I(\eta)$ -see (19)- is evaluated using an adaptive Simpson's rule, starting from 17 equidistant evaluation points. For all components of θ, the minimum and maximum values are set at -30 and 30 respectively.

Table 3: Results concerning the trimodal distribution

	APMH	APIS	MH	IS	true
mean(θ_1)	-5.91	-5.80	8.00	-12.00	-5.33
mean(θ_2)	-5.85	-5.49	8.00	-12.01	-5.33
mean(θ_7)	1.33	1.13	8.00	-11.99	1.33
mean(θ_8)	1.42	1.23	8.00	-12.00	1.33
stdev(θ_1)	10.19	10.23	3.99	4.00	10.24
stdev(θ_2)	10.31	10.22	4.00	4.01	10.24
stdev(θ_7)	10.36	10.32	4.00	4.00	10.24
stdev(θ_8)	10.29	10.34	4.00	4.00	10.24
corr(θ_1,θ_2)	0.85	0.85	0.00	0.00	0.85
corr(θ_6,θ_7)	0.42	0.43	0.00	0.00	0.42
corr(θ_7,θ_8)	0.84	0.85	0.00	0.00	0.85
Mahalanobis	0.02	0.03	0.00	0.00	

We emphasize that the numerical results reported below depend on the design of the experiments and on the initial values. Different initial values (seed of the random number generator and starting values of the location and scale) may give different numerical results. However, when the sample size is increased, the reported results are robust.

Table 2 and Table 3 contain the results of the experiments. A subset of the estimated means, standard deviations and correlations between subsequent components are reported (the results for the other elements of θ are very similar). True values for the estimated moments can be found in the last column of the tables. Further, the Mahalanobis distance (see Appendix 2) at round 8, indicating convergence if close to 0, is shown.

Both Table 2 and Table 3 show a superior performance of APMH and APIS over MH and IS. Let us consider the unimodal high-correlation distribution first. It is seen that the

two polar algorithms have succeeded in obtaining good estimates, although a minor improvement on the location is still possible. On the other hand, MH and IS have completely lost the track. The location is incorrect and most of the correlations are much too low. Moreover, since the scale of the updated normal candidate distribution is too small by far, as all standard deviations are smaller than 1 while they should equal 5, there is no perspective of improvement.

The results concerning the trimodal distribution are even more striking. Although APMH and APIS have not completely converged yet, their estimates are reasonably accurate. This is quite in contrast with MH and IS. Both algorithms miss two of the three modes. MH is completely stuck in the third mode, whereas IS is not able to escape from the first one. For illustrative purposes, Figure 3 and Figure 4 display estimated marginal densities of the trimodal distribution.

As the previous examples are perhaps too favourable to the adaptive polar algorithms, we consider a wider range of experiments which differ in the number of dimensions and modes of the target density and in the distance between the modes. The object of the exercise is to see which algorithms are able to sample from the target distribution without missing out a mode, and which algorithms fail. In this set of experiments, we also consider the griddy-Gibbs (GG) sampler as a competitor, see Bauwens and Lubrano (1998) for a presentation of the GG sampler.

The target density of dimension k and distance parameter r has k modes at multiples $re_1 = (r, 0, .., 0)', .., re_k = (0, .., 0, r)'$ of the unit vectors. It is the following normal mixture:

$$\theta \sim \sum_{i=1}^{k} k^{-1} \mathcal{N}(re_i, I_k). \tag{27}$$

It is unimodal if $k = 1$, but as k increases, the number of modes increases, and also the modes are ever harder to distinguish from the mass of the density around the origin.

Sampling is done in an automatic procedure, sampling 100,000 drawings which are distributed over 10,000 directions for the APMH and APIS algorithms. When using a Metropolis-Hastings step, with APMH and with MH itself, both accepted and unaccepted drawings are counted. The initial location estimate is the origin, with initial covariance matrix extremely vague, $\Sigma_{\text{init}} = 100 I_k$. Location and scale estimates are updated after sampling, and when the Mahalanobis distance improves by more than 50%, sampling is repeated.[5]

Before commenting on a set of sampling results in Table 4, first look at a selection of choices of k and r, in Figure 5. The panels display the marginal densities of the elements of θ for four special cases. At a dimension of $k = 4$ and with a short distance between the modes, with $r = 4$, the APIS and GG samples are displayed in the top row of the figure. Both algorithms (like APMH, MH and IS) have no trouble in locating both modes, and all marginal densities are recovered neatly.

At a higher dimension, with more modes, getting the correct distributions is harder. The bottom row displays results for the APMH and MH algorithms, at $k = 8$ and $r = 16$. The APMH results show that for all elements of θ, the bimodality is found by the sampler, though the estimates are not too precise. Indeed, if sampling is continued, updating the location and scale more often, full convergence to the correct densities does not seem to be guaranteed.

---

[5]Such an automated stopping rule works reasonably well over the experiments of this setup. In practice, it is advisable to monitor convergence more carefully.

Figure 3: Univariate marginal densities for the trimodal distribution. The plots in the first column result from APMH, those in the second column result from APIS.
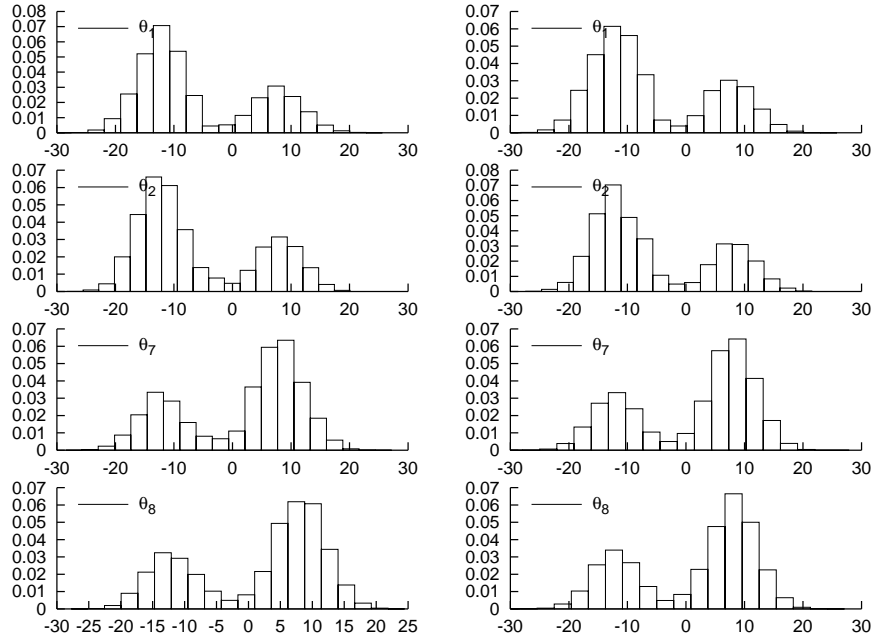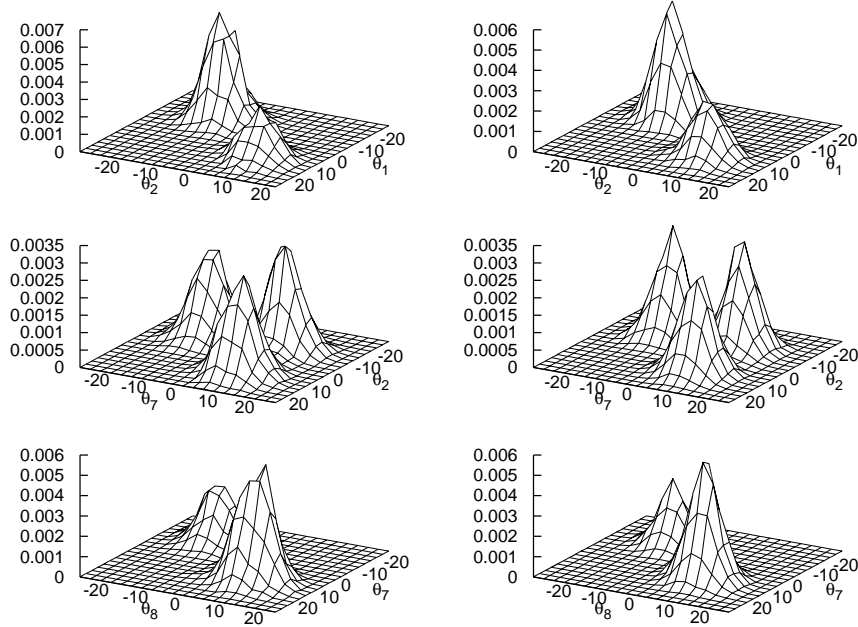

Figure 4: Bivariate marginal densities for the trimodal distribution. The plots in the first column result from APMH, those in the second column result from APIS.
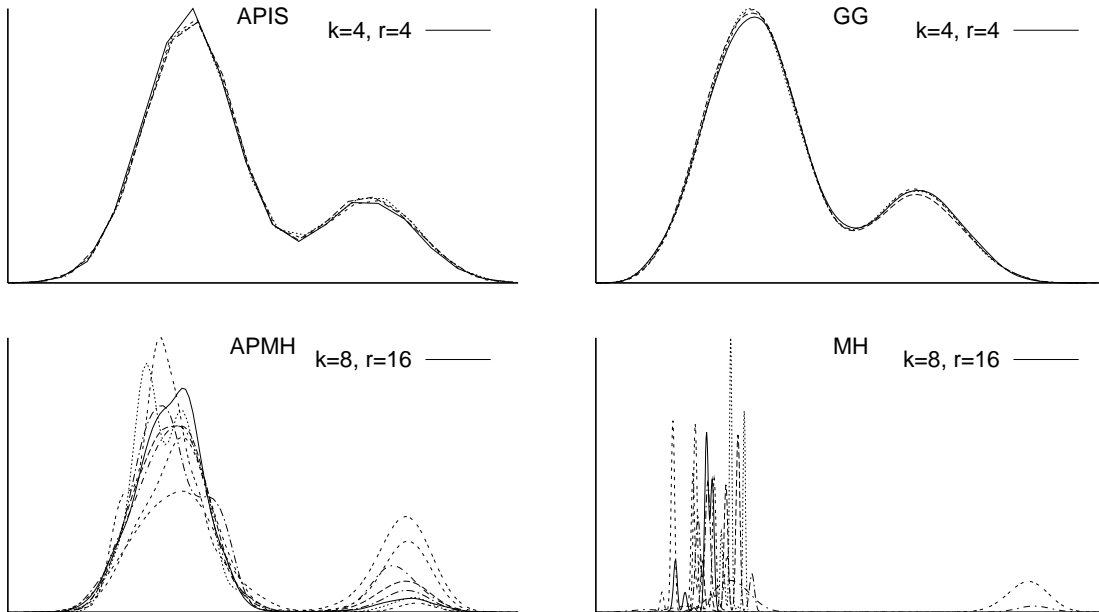
Figure 5: A selection of sampled marginal densities, for APIS ($k = 4, r = 4$, top left), GG ($k = 4, r = 4$, top right), APMH ($k = 8, r = 16$, bottom left) and MH ($k = 8, r = 16$, bottom right)

With the MH results on the right, only for two of the 8 elements of θ the bimodality is recognized. More detailed sampling results (not reported here) show that the MH algorithm is not able to get a good initial sample from the target density in order to update the location and scale of the candidate density. The final acceptance rate lies below 1%, a clear sign of non-convergence. Making a larger sampling effort does not help in finding the multimodality of the target, but rather leads to convergence around the single mode at the origin.

Each block of Table 4 corresponds to a target density (a k, r combination). The first row of each block reports the Euclidian length of the true mean vector μ of the target density (in column 2) and a summary measure of relative error on the estimated mean vector, for each sampling method (in the next columns). This measure is the length of the difference between the vectors of estimated and true means, divided by the length of the true mean vector (say $|\overline{x} - \mu| / |\mu|$ where $\overline{x}$ is the estimated mean vector and $|a|$ denotes the length of a). The second row gives the rank of each sampling method according to that criterion.

From the results in the table, we see that for the small distance of $r = 4$, all algorithms give a small relative error, with the exception of MH and IS when $k = 8$. Not surprisingly, when increasing r, the relative errors increase in all cases. When the modes lie at a distance of $r = 16$, it really is hard to retrieve the modes. If there are only four modes, most algorithms still deliver reasonable precision; with higher dimensional problems, even APMH and APIS

15

Table 4: Comparison of sampling methods on normal mixtures

	Model	APMH	APIS	MH	GG	IS
			$r = 4$			
$k = 4$	2.00	0.02	0.03	0.03	0.02	0.09
		1	3	3	1	5
$k = 8$	1.41	0.16	0.06	1.87	0.01	2.64
		3	2	4	1	5
			$r = 8$			
$k = 4$	4.00	0.06	0.08	0.06	0.74	0.13
		1	3	1	4	5
$k = 8$	2.83	0.46	0.58	2.68	0.86	2.08
		1	2	5	3	4
			$r = 16$			
$k = 4$	8.00	0.08	0.13	0.15	1.74	0.18
		1	2	3	5	4
$k = 8$	5.66	0.80	1.89	2.21	2.66	2.79
		1	2	3	4	5
Global ranking		1	2	3	4	5

For each k, r combination, the first row gives $|\mu|$ (in the column 'Model'), the length of the vector of true means of the target density. To the right of it, for each sampling method one finds $|\overline{x} - \mu|/|\mu|$, the ratio of the length of $\overline{x} - \mu$ to $|\mu|$, where $\overline{x}$ is the estimated mean vector. In the second row, one finds the rank of each method according to the relative error in the first row. The target density is defined in eq (27).

start to miss the correct mean.

The last row of Table 4 gives a global ranking across a range of experiments which combine $r = 2, 4, 8, 12, 16$ with $k = 4, 6, 8$ (the table reports the results only for a subset of these combinations). The ranking is based on the addition of the ranks across all experiments. The conclusion is that for this type of multimodal target densities, the adaptive polar algorithms are the best methods.[6]

4   Application to econometric models

In this section, a set of models that we encountered in practice are used to illustrate the versatility of the polar algorithms. We do not claim that the algorithms are necessarily the most efficient ones for the analyzed models and data, since a careful analysis of the posterior density may help to design a more efficient algorithm (actually, the polar algorithms can also be useful in this perspective). For completeness, we also compare APMH and APIS to MH, GG, and IS.

---

[6]These findings are similar to the ones in Monahan (2001), who also notes that with the method of spherical-radial integration, which is similar to the MIXIN algorithm of Van Dijk, Kloek, and Boender (1985), it is hard to find modes far away from the origin.

4.1 Scale contamination

In Justel and Peña (1996), the authors investigate a data set from Brownlee (1965, pp. 491–500) concerning the oxidation of ammonia to nitric acid in a plant. The data set incorporates 21 daily observations on four variables. The uptake of nutrients y by the plant is related to the amount of air flow, the temperature of the cooling water, and the concentration of acids (taken up into the vector x). In several investigations it was found that several observations might be classified as outliers, and therefore care should be taken in the analysis to allow for this. In a regression setting, it is sufficient to allow for scale contamination, as in the model

$$y_i = \beta' x_i + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma_i^2)$$
$$\sigma_i = \begin{cases} \sigma & \text{with probability } 1 - \alpha \\ \kappa\sigma & \text{with probability } \alpha \end{cases}$$

with the restrictions $0 \leq \alpha \leq 1$ and $\kappa \geq 1$. The priors for κ, α, and each element of β are taken uniform on the intervals [1, 10], [0, 1], and [-30, 30], respectively. The conditional prior $\pi(\sigma|\kappa, \alpha)$ is proportional to $[(1 - \alpha)\sigma + \alpha\kappa\sigma]^{-1}$ with σ in the region (0, 10).

The sampling algorithms were initialized by searching for the mode of the posterior density. A border solution with $\alpha = 0$ was found, indicating that the scale contamination is not immediately apparent from the posterior density. However, with the result of the optimization procedure only serving as a starting point, the algorithms were started to find the posterior density of the parameters. Table 5 reports the means and standard deviations of the final samples. The lower panel contains statistics related to the final sample and other information on each algorithm (see Appendix 2 for details on the sampling setup and on the reported statistics).

The acceptance rate of directions η for APMH of 0.59 compares favourably to the acceptance rate for MH of 0.30. The result is a lower correlation in the sample of APMH (see $\rho_{\max}$), while GG (griddy-Gibbs) displays a tremendous correlation of 0.994 between successive drawings. The final value of the Mahalanobis measure, measuring the change in location with respect to the previous candidate, is still rather high for MH and IS.

The estimated mean of κ using MH differs slightly from the estimates using the other algorithms. Looking at the posterior density of α and κ in Figure 6, it seems that MH stumbles over the long right tail of the density for κ (top middle panel), and as a consequence, it underestimates the posterior standard deviations of all the parameters. The other algorithms find the heavy tail of the posterior density of κ.

The posterior density of α displays a slight bimodality: the mass for interior values of $0 < \alpha < 1$ indicates that some evidence is found for the scale contamination, but both corner solutions of $\alpha = 0$ and $\alpha = 1$ receive positive mass as well. In the case $\alpha \approx 0$, the value of κ is not defined, which creates the long right tail of the density. This effect is clearly apparent from the plots of the joint posterior density of α and κ, in the lower panels of Figure 6.

4.2 Predictability in an exchange rate series

In Bos, Mahieu, and Van Dijk (2000) (and more elaborately in Bos (2001)) the daily German mark-USA dollar exchange rate series[7] is taken as the basis for providing a decision whether

---

[7]Daily data were extracted from Datastream, for the period 1/1/1982-31/12/1997.

Table 5: Posterior results for scale contamination model

Method	Initialization		APMH		APIS		MH		GG		IS	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Air Flow	0.796	(0.22)	0.804	(0.20)	0.824	(0.20)	0.795	(0.18)	0.805	(0.17)	0.795	(0.18)
Water Temp	1.110	(0.22)	1.032	(0.57)	0.965	(0.57)	1.073	(0.49)	1.079	(0.54)	1.051	(0.51)
Acid Conc	−0.624	(0.22)	−0.611	(0.09)	−0.609	(0.09)	−0.615	(0.09)	−0.623	(0.10)	−0.611	(0.09)
σ	3.733	(0.22)	3.031	(1.36)	3.003	(1.41)	3.407	(0.95)	3.051	(1.34)	3.197	(1.29)
κ	1.477	(0.22)	3.430	(2.43)	3.474	(2.37)	2.261	(1.18)	3.481	(2.46)	3.432	(2.26)
α	0.000	(0.22)	0.432	(0.33)	0.449	(0.33)	0.389	(0.30)	0.431	(0.33)	0.418	(0.34)
α_η or α_θ			0.588				0.298					
ρ_{max}			0.665				0.882		0.994			
Mahalabonis			0.023		0.027		0.251				0.150	
Nit			9		9		8		1		10	
Fn ($\times 1000$)			316		183		34		6246		10	

Reported are the means (μ), standard deviations (σ), acceptance rate (α_η or α_θ), maximal serial correlation (ρ_{max}), and Mahalanobis distance (all these statistics pertain to the sample generated after the last update of the candidate, and some are not relevant in some cases). The number of updates of the candidate in adaptive algorithms (Nit), and the total number of function evaluations (Fn) are also reported (see Appendix 2 for details).
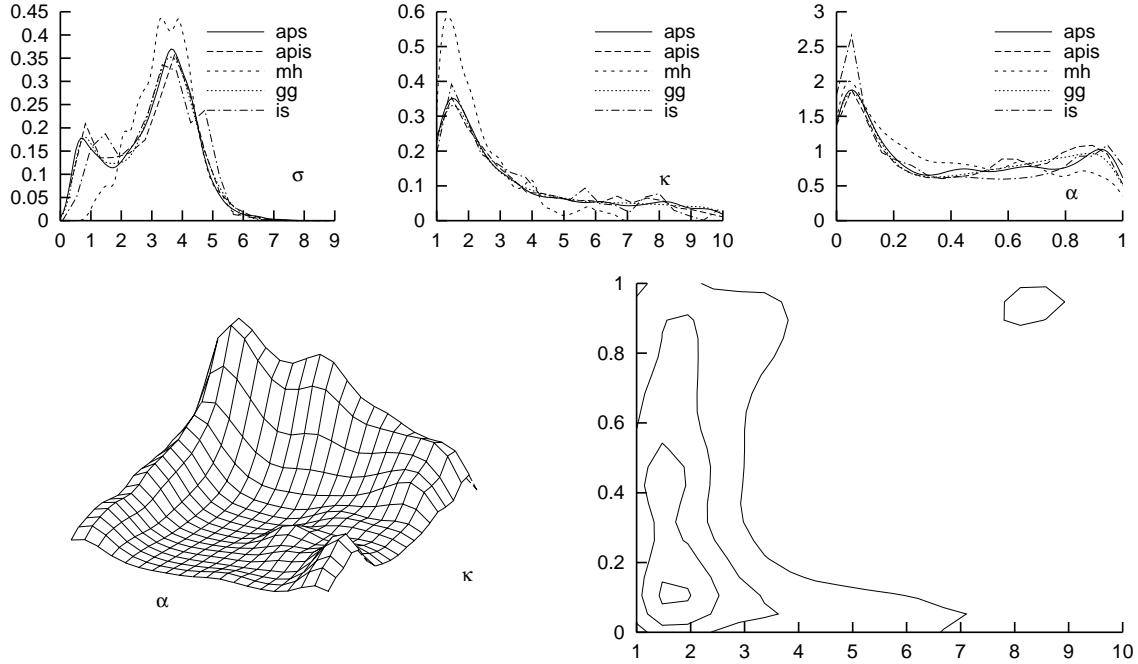
Figure 6: Posterior densities of α and κ for scale contamination model

or not to hedge the currency risk. For an optimal decision, one wants to make use of all possible information on the predictability of the series. Essentially, the model used is

$$y_t - c = \phi(y_{t-1} - c) + \epsilon_t + \theta\epsilon_{t-1}$$

$$\epsilon_t \sim t\left(0, \sqrt{\frac{\nu - 2}{\nu}}h_t, \nu\right)$$

$$h_t = \delta h_{t-1} + (1 - \delta - \alpha) + \alpha\epsilon_{t-1}^2$$

where $y_t = 100(\log Y_t - \log Y_{t-1})$ is the percentage change in the exchange rate Y_t and ϵ_t is Student-t distributed with expectation 0, variance h_t and ν (> 2) degrees of freedom. For the priors, we choose $c \sim N(0, 0.0004)$, ϕ and $-\theta \sim N(0.8, 0.04)$, δ and α jointly uniform on the region ($0 < \delta$, $0 < \alpha < 1$, $0 < \delta + \alpha < 1$), $\sigma \sim$ inverted-gamma-1 with shape parameter 2.5 and scale parameter 1.333 (see Bauwens, Lubrano, and Richard (1999) for the definition), and ν proportional to a Cauchy density truncated to the region $\nu > 2$ (see Bos, Mahieu, and Van Dijk (2000) for details). Note how a priori we do expect non-zero values for ϕ and θ, but with very near root cancellation, inducing the type of slight autocorrelation seen in exchange rate series.

Results of the simulations are presented in Table 6 and Figure 7. All samplers do converge, provided many updates are performed. The curse of dimensionality hitting the griddy-Gibbs sampler is clear from the number of function evaluations: the sampler needed 6 million evaluations for constructing a sample of 10000 drawings. Again, the maximal first-order serial correlation is lowest for APMH, with MH coming in at the second place and the GG in the last position with very strong autocorrelation between successive drawings. The slow mixing behaviour of the latter sampler is apparent from Figure 7: The parameters ϕ and θ are

19

Table 6: Posterior results for ARMA-GARCH-Student t model

Method	Initialization		APMH		APIS		MH		GG		IS	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
c	0.006	(0.01)	0.006	(0.01)	0.006	(0.01)	0.006	(0.01)	0.006	(0.01)	0.006	(0.01)
ϕ	0.844	(0.14)	0.770	(0.12)	0.770	(0.12)	0.771	(0.12)	0.759	(0.11)	0.770	(0.12)
θ	−0.830	(0.14)	−0.755	(0.12)	−0.755	(0.12)	−0.756	(0.12)	−0.745	(0.12)	−0.755	(0.12)
σ	0.741	(0.06)	0.758	(0.07)	0.761	(0.07)	0.757	(0.07)	0.757	(0.07)	0.757	(0.07)
δ	0.919	(0.01)	0.916	(0.01)	0.916	(0.01)	0.916	(0.01)	0.916	(0.01)	0.916	(0.01)
α	0.063	(0.01)	0.065	(0.01)	0.065	(0.01)	0.065	(0.01)	0.065	(0.01)	0.065	(0.01)
ν	5.214	(0.43)	5.282	(0.44)	5.289	(0.46)	5.274	(0.44)	5.285	(0.44)	5.291	(0.46)
α_η or α_θ			0.760				0.447					
$\rho_{\max}$			0.383				0.639		0.993			
Mahalabonis			0.018		0.014		0.008				0.008	
Nit			9		10		11		1		10	
Fn ($\times 1000$)	199				152		22		6122		10	

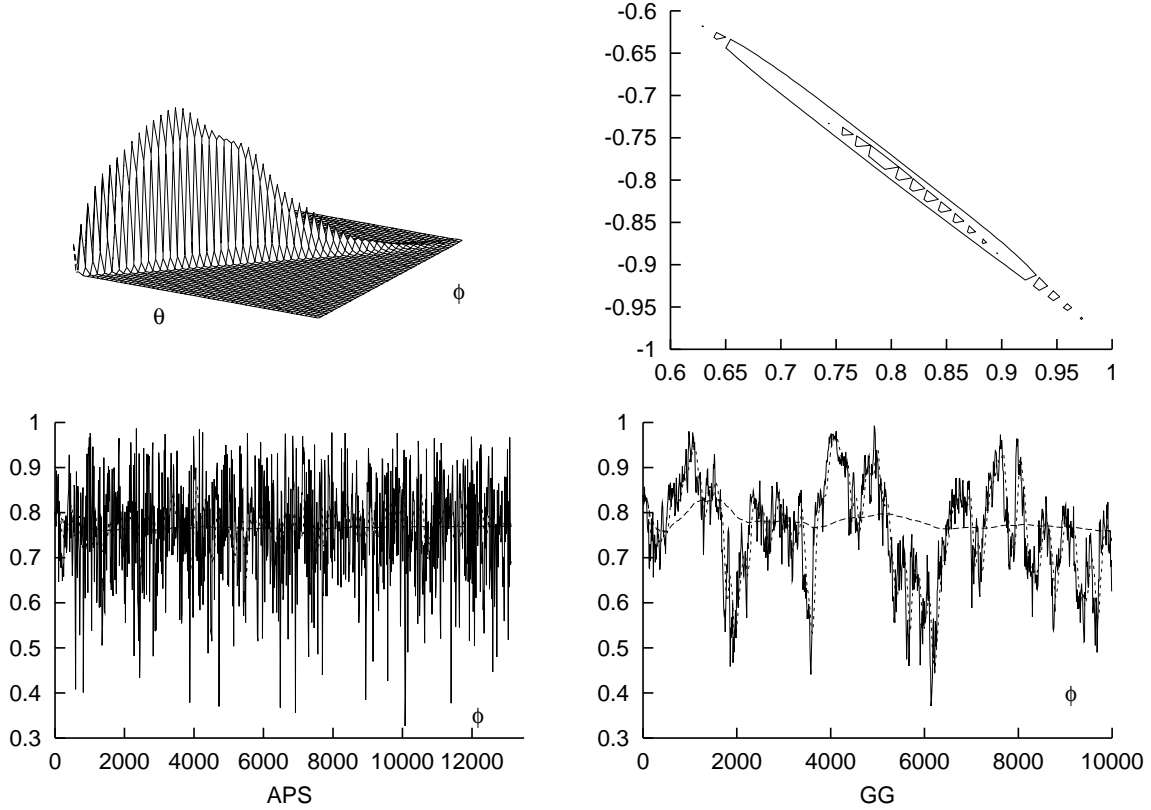See Table 5 for an explanation of the entries in this table.

Figure 7: Bivariate posterior density of ϕ and θ for the ARMA-GARCH-Student-t model, and plots of the sampled values and cumulative/running means for APMH and GG

strongly correlated along the line $\phi = -\theta$ of root cancellation, which implies that sampling from e.g. $\phi | \theta, \ldots$ leaves only little freedom for obtaining a radically different value of ϕ. In the bottom right panel this is seen clearly: ϕ moves around the parameter space only very slowly, while APMH does not have trouble at all (see the bottom left panel).

4.3 A mixture model for the USA GNP growth rate

In models for the growth rate of the gross national product, great advances have been made by allowing for separate regimes in periods of recession and expansion. However, these models give rise to difficulties with respect to convergence of sampling methods due to multiple modes. We make use of APMH and APIS, together with the more usual sampling methods, on three implementations of the mixture model. For an other recent method we refer to Frühwirth-Schnatter (2001). We consider as data the growth rate of the real GNP of the USA (see Figure 8).[8]

---

[8]Source: Gross national product in billions of dollars, seasonally adjusted at an annual rate, period 1959:I until 2001:I, from Economagic. The growth rate is measured as 100 times the first difference of the logarithm.
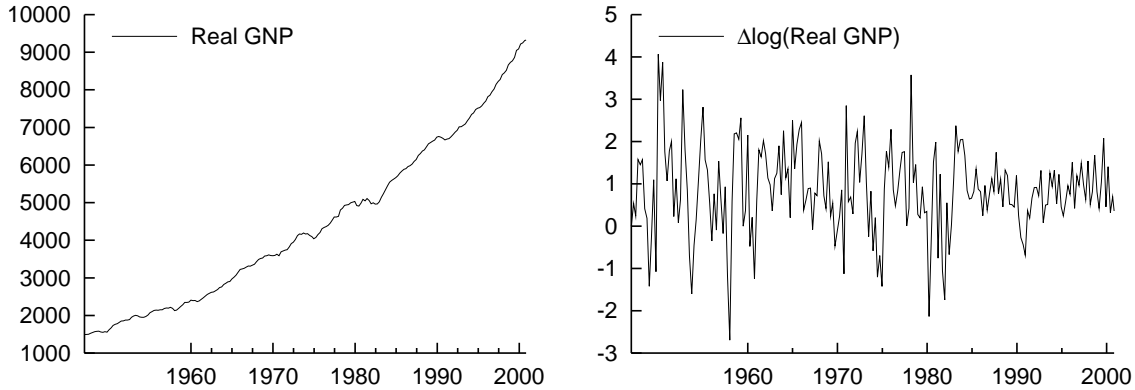
Figure 8: Real USA GNP, 1959:I–2001:I

A bivariate mixture

The most simple model, useful to show the occurrence of multiple modes in models of this kind, is a bivariate mixture model. It allows the growth rate to display two distinct mean levels, modelled as

$$y_t = \epsilon_t + \begin{cases} \beta_1 & \text{with probability } p \\ \beta_2 & \text{with probability } 1 - p \end{cases} \tag{28}$$

with $\epsilon_t \sim N(0, \sigma^2)$. For identification we assume that $\beta_1 < \beta_2$, such that the first regime is the low growth regime. Priors on the parameters β_1, β_2 and p are taken uniform, for σ we use the uninformative prior $\pi(\sigma) \propto 1/\sigma$, with σ restricted to a finite range to ensure existence of all conditional posterior densities.

Results are in the top panel of Table 7. IS results in quite different posterior moments for p and β_1. In Figure 9 this difference comes to light. The importance sampler seems not to have converged, displaying several spikes in the density plot of p (and less strongly so for the other parameters). This is caused by sampling in the tail of an importance density which did not fit the posterior very well.

The results of the other samplers correspond quite well, especially GG and APMH. However, the GG algorithm took 12.5 million function evaluations, whereas the APMH (and APIS) needed only 500 thousands. MH is even more efficient in terms of function evaluations that are needed.

The probability of being in the low growth regime is estimated, rather imprecisely, to be about one third for this data set. All sampling methods recognize that there is a probability that $\beta_1 \approx \beta_2 \approx 1$, with p badly identified (its posterior density has a long, flat, right tail). IS is not able to sample sufficiently from this region, leading to a wrong estimate of the posterior density of β_1.

Note the small mode in the posterior density of p at about 0.9. This value corresponds to a situation where most observations come from a 'low' growth regime (of around 1% per quarter), with a small probability of observing a high growth rate (greater than 1%). The importance of this bimodality in p could be investigated by allowing for a third regime, thus allowing for periods of negative, low, and high growth.

22

Table 7: Posterior results for the mixture models for GNP growth

Method	Initialization		APMH		APIS		MH		GG		IS	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Using the two-regime mixture model												
p	0.041	(0.04)	0.355	(0.35)	0.326	(0.32)	0.314	(0.31)	0.355	(0.34)	0.207	(0.19)
σ	0.919	(0.06)	0.959	(0.07)	0.963	(0.07)	0.984	(0.08)	0.965	(0.07)	0.927	(0.07)
β_1	-1.243	(0.62)	-0.098	(0.90)	-0.099	(0.89)	-0.162	(0.99)	-0.075	(0.91)	-0.308	(0.91)
β_2	0.929	(0.08)	1.220	(0.59)	1.083	(0.31)	1.079	(0.26)	1.178	(0.54)	1.064	(0.21)
α_η or α_φ			0.446				0.184					
$\rho_{\max}$			0.969				0.928		0.610		0.413	
Mahalanobis			0.101		0.014		0.250					
Fn ($\times 1000$)			506		517		347		12578		22	
nIt			6		8		9		1		8	
$\log(\mathrm{ML}_{\mathrm{M2}})$			-316.19		-316.40		-316.22		-316.53		-324.84	
Using the three-regime mixture model												
p_1	0.333	(1.00)	0.299	(0.26)	0.351	(0.26)	0.309	(0.27)	0.307	(0.27)	0.053	(0.04)
p_2	0.333	(1.00)	0.462	(0.29)	0.425	(0.24)	0.443	(0.27)	0.435	(0.28)	0.678	(0.24)
σ	1.017	(1.00)	1.461	(0.11)	1.446	(0.11)	1.454	(0.10)	1.455	(0.10)	1.403	(0.08)
β_1	-0.203	(1.00)	0.524	(0.34)	0.566	(0.35)	0.484	(0.43)	0.450	(0.56)	0.172	(0.43)
β_2	0.815	(1.00)	0.849	(0.18)	0.883	(0.17)	0.868	(0.19)	0.855	(0.21)	0.762	(0.14)
β_3	1.877	(1.00)	1.311	(0.53)	1.627	(0.91)	1.387	(0.76)	1.384	(0.75)	1.087	(0.24)
α_η or α_φ			0.413				0.124					
$\rho_{\max}$			0.876				0.867		0.446		0.923	
Mahalanobis			0.016		0.230		0.049					
Fn ($\times 1000$)			2076		634		1794		23950		13	
nIt			9		10		8		1		7	
$\log(\mathrm{ML}_{\mathrm{M3}})$			-300.36		.		-298.67		-298.89		-301.80	
$\log(\mathrm{BF}_{\mathrm{M3\text{-}M2}})$			15.83		11.38		17.55		17.64		23.04	
Using the two-regime AR(1) mixture model												
p	0.776	(9.52)	0.410	(0.30)	0.308	(0.25)	0.472	(0.32)	0.580	(0.36)	0.531	(0.25)
σ	0.901	(8.15)	0.901	(0.06)	0.901	(0.04)	0.900	(0.05)	0.897	(0.06)	0.900	(0.05)
β_1	0.275	(8.15)	-0.098	(0.51)	-0.212	(0.36)	0.007	(0.47)	0.077	(0.51)	0.102	(0.32)
β_2	1.325	(8.15)	0.958	(0.43)	0.888	(0.18)	1.108	(0.61)	1.631	(1.15)	1.024	(0.22)
ϕ_1	0.485	(0.91)	0.645	(0.21)	0.576	(0.17)	0.604	(0.22)	0.557	(0.21)	0.586	(0.21)
ϕ_2	0	(15.27)	0.197	(0.13)	0.181	(0.10)	0.191	(0.11)	0.228	(0.17)	0.140	(0.07)
α_η or α_φ			0.357				0.129					
$\rho_{\max}$			0.928				0.942		0.832		1.828	
Mahalanobis			0.058		0.963		0.210					
Fn ($\times 1000$)			1232		289		1250		19122		32	
nIt			8		8		7		1		8	
$\log(\mathrm{ML}_{\mathrm{AR}})$			-302.81		-313.01		-301.75		-303.10		-305.75	
$\log(\mathrm{BF}_{\mathrm{AR\text{-}M2}})$			13.38		11.38		14.47		13.43		19.09	

See Table 5 for an explanation of the entries in this table.
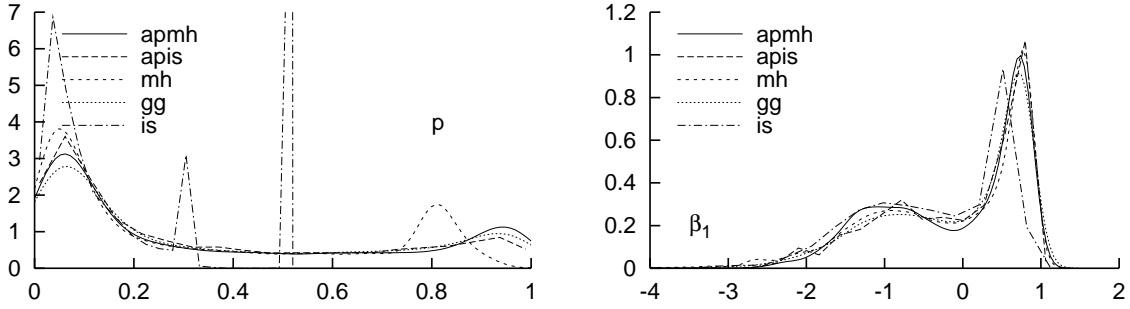
Figure 9: Posterior densities of p and β_1 in the two-regime mixture model for GNP growth

Adding a third regime

The bimodality of the posterior density of β_1 in the right panel of Figure 9 could indicate a misspecification of the model. With a three regime model, allowing for periods of recession, normal growth and rapid expansion, possibly a clearer distinction between regimes could be found. We model this as

$$
y_t = \epsilon_t + \begin{cases} \beta_1 & \text{with probability } p_1 \\ \beta_2 & \text{with probability } p_2 \\ \beta_3 & \text{with probability } 1 - p_2 - p_1. \end{cases} \tag{29}
$$

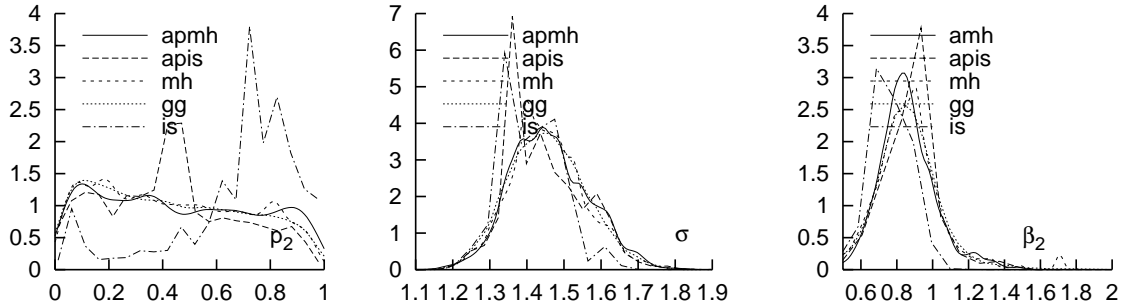For identification, we use the the restriction $\beta_1 < \beta_2 < \beta_3$.

Figure 10: Posterior densities of p_2, σ and β_2 in the three-regime mixture model for GNP growth

The second panel of Table 7 gives the results for this enlarged model. Figure 10 displays the marginal posterior densities of p_2, σ, and β_2. From the leftmost panel it is clear that the methods based on importance sampling (APIS and IS) seemingly do not give correct results on this model–data combination. The other three methods agree, mostly on the message that it is difficult to discriminate between the regimes. The data set contains only 42 years of quarterly data, which has to be spread over the three regimes. Therefore, it is no wonder that it is hard to identify clearly three regimes.

Contrasting the two and three regime models is best done by comparing the marginal likelihoods of the model (see Aitkin 1991, Kass and Raftery 1995). Rows indicated with

24

log(ML) in Table 7 report the logarithm of the marginal likelihood of the model.[9] The line indicated with log(BF) provides the logarithm of the Bayes factor, which is the difference between the log-marginal likelihoods. According to the classification of Kass and Raftery (1995), a value of about 16 can be considered 'very strong evidence' in favour of the model containing three regimes.

Adding time dependence in the bivariate mixture

The mixture models have no time dependence in them. The model which Frühwirth-Schnatter (2001) uses is a mixture model with two regimes, and with AR(1) within both regimes:[10]

$$
y_t = \epsilon_t + \begin{cases} \beta_1 + \phi_1 y_{t-1} & \text{with probability } p \\ \beta_2 + \phi_2 y_{t-1} & \text{with probability } 1 - p \end{cases}
\tag{30}
$$

The third panel in Table 7 shows the posterior moments and other statistics for this third model.
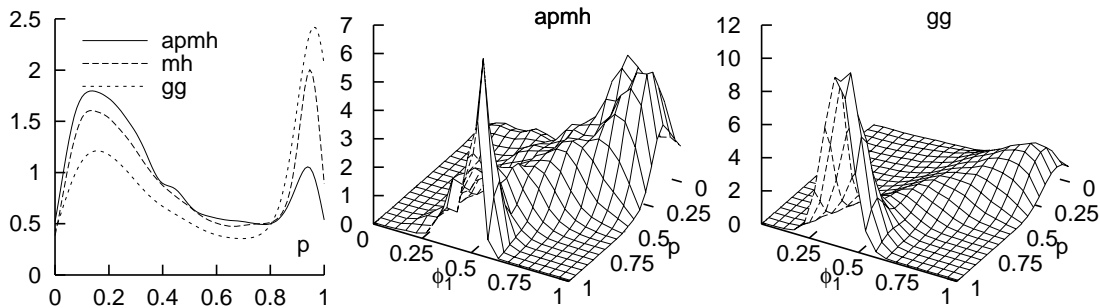

Figure 11: Posterior density of the parameter p in the bivariate mixture with AR-components-model, for the APMH, MH and GG samplers, plotted against the density of ϕ_1 for APMH and GG.

For this model, the simulation methods do not agree closely as far as the posterior moments are concerned. The general impression from the plot of the estimated posterior density of p (in the leftmost panel of Figure 11) is that the sampling methods judge the relative importance of both regimes differently, with APMH placing less weight on the second regime.

The fifteen-fold larger computational effort for the Griddy Gibbs sampler leads to a smoother estimate of the marginal joint posterior density of p and ϕ_1, than for the APMH sampler. The dependence structure between these two parameters however looks the same.

The lower lines of Table 7 report the log-marginal likelihood and the log-Bayes factor. Evidence is strong that the two regime model with AR-components is better in explaining the data than the two-regime static model presented before, but the three regime static model fits the data even better.

---

[9]The marginal likelihood is calculated as the difference between the log-posterior density at the posterior mean of the sample and the logarithm of an approximating kernel density at the same location, see Bos (2002) for a comparison of this and other methods . This measure is sufficiently stable in models of low dimension, and is easily computed.

[10]Actually, Frühwirth-Schnatter (2001) also allows for the variances to differ between regimes.

4.4 A TAR model for the growth rate of USA industrial production

In growth rate models of industrial production, a common assumption is to allow for different regimes depending on the state of the economy. One of the structures arranging for such behavior is the smooth transition autoregressive (or STAR) model, which presumes that the growth rate follows a weighted average of two autoregressive processes, with weights depending on the current state of the economy.

The model is

$$y_t = \phi_A^{(p)'} y_t^{(p)} F(s_t; \gamma, c) + \phi_B^{(p)'} y_t^{(p)} (1 - F(s_t; \gamma, c)) + \epsilon_t \tag{31}$$

with $\phi_i^{(p)} = (\phi_{i0}, \phi_{i1}, .., \phi_{ip})'$ $(i = A, B)$ the AR parameters of regime i, and $y_t^{(p)} = (1, y_{t-1}, .., y_{t-p})'$. The error terms u_t are assumed to be white noise, with mean 0 and variance σ^2.

Instead of ϕ_{i0}, the constant in the AR equation, a parameter $\mu_{i0} = \phi_{i0}/(1 - \sum_j \phi_{ij})$ is used. This parameter μ_{i0} displays less correlation with the AR parameters than ϕ_{i0} itself. Also, it is more straightforward to devise a sensible prior for the mean of the regime, than for the constant parameter in the AR polynomial.

For the weighting or transition function $F(s_t; \gamma, c)$ a common choice is the logistic function

$$F(s_t; \gamma, c) = \frac{1}{1 + e^{-\gamma(s_t - c)}}, \text{ with } \gamma > 0. \tag{32}$$

The parameter γ is restricted to be positive, otherwise it would not be identified. The variable s_t triggers the transition from one regime to the other. Possible implementations use y_{t-d} for a fixed number of lags d, or Δy_{t-d}.

A preliminary analysis indicated that the transition from one regime to the other is swift, such that the threshold autoregressive (TAR) model (with $\gamma \to \infty$) works better. As the threshold variable, the sum of the growth rate over the previous three months is used.

As data, we use the industrial production index in the USA as provided by the Federal Reserve Board, over the period 1961:1–1998:12, see Figure 12.[11] Hence, y_t in (31) is the first difference of the logarithm of the industrial production index, multiplied by 100.

Parametrization and the prior

As growth of the industrial production is a stable process (in the sense that explosive behaviour is very unlikely), we would like to put most of the prior weight for the parameters on the stationary region. However, for (S)TAR models of autoregressive order $p > 1$, little is known about the conditions on the parameters to attain stationarity (see Chan, Petrucelli, Tong, and Woodford 1985, Enders and Granger 1998). Therefore, we 'encourage' each of the regimes to have stationary roots in the lag polynomial. This is implemented by using a prior $\pi(r_{i1}, .., r_{ip}) \sim \mathcal{N}(0, 0.45 \times \mathcal{I}_p)$ on the inverse roots r_{ij} of the polynomial, such that each inverse root lies within the unit circle with probability 0.86.

Each of the regimes has a separate mean governed by the parameter ϕ_{i0} in relation to the other autoregressive parameters $\phi_{i1}, .., \phi_{ip}$. Instead of putting a prior on the constant in the

---

[11] The series is series number B50001 from the Federal Reserve Board, on Industrial Production: Market Groups and Industry Groups, group G17, seasonally adjusted.
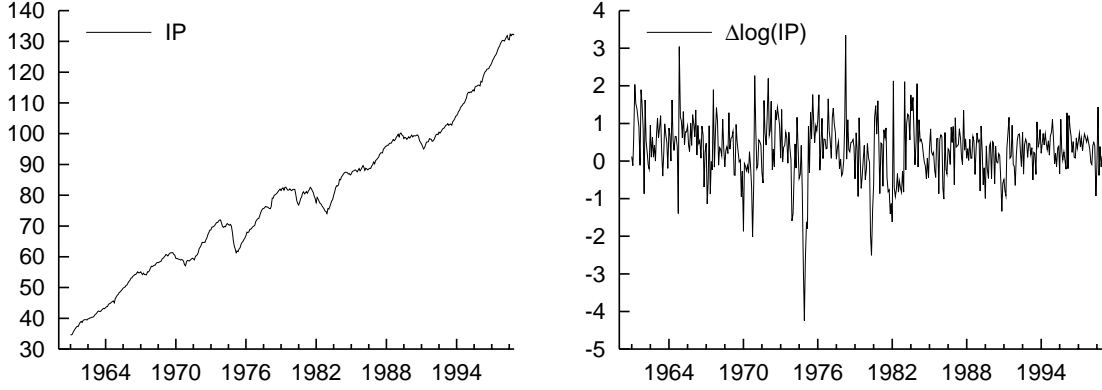
26

Figure 12: Industrial production in the USA, 1961:1–1998:12

AR-polynomial directly, we instead put it on

$$
\mu_{i0} = \begin{cases} \frac{\phi_{i0}}{1 - \phi_{i1} - .. - \phi_{ip}} & \text{if the regime is stable} \\ \phi_{i0} & \text{otherwise.} \end{cases}
$$
(33)

The prior on μ_{i0} is

$$
\pi(\mu_{i0}) \sim \begin{cases} \mathcal{N}(3.5, 1.4^2) & \text{if the regime is stable} \\ \mathcal{N}(0.5, 0.5^2) & \text{otherwise.} \end{cases}
$$
(34)

The prior for c is data-based, as the meaning of the parameter depends entirely on the underlying trigger variable. We use a normal prior, with mean equal to the mean of the trigger s_t and corresponding variance. Note that with a TAR model, the precise value of c is not important, it is its relation to the order statistics of the trigger variable s_t which counts. Instead of this continuous distribution for c, a second option would be to use a discrete distribution with support at all the order statistics of s_t.

Sampling results

Numerical results are given in Table 8 for a TAR(3) model. From the estimated standard deviations it can be seen that the parameters for regime A are harder to estimate than those of regime B. Indeed, the threshold c of around -2.4 implies that less than 5% of the observations are classified as coming from the first regime, leaving most observations to estimate the parameters in regime B.

The AR-parameters of the second regime indicate stable behaviour, with an average growth rate of about 0.3% per quarter (which corresponds to the mean growth of the series itself). Only the griddy-Gibbs results differ strongly: close examination of the posterior density of the parameters for the griddy-Gibbs algorithm shows that the algorithm continues to select values of $c \approx 6$, which would place all observations in the A-regime. This seems to indicate non-convergence of the Gibbs algorithm, with the algorithm getting stuck in a 'corner-solution'.

27

Table 8: Posterior results for TAR model

Method	Initialization		APMH		APIS		MH		GG		IS	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
μ_A	3.669	(1.09)	2.517	(1.89)	2.367	(1.83)	2.601	(1.84)	1.527	(1.65)	3.257	(1.70)
ϕ_{1A}	0.860	(0.03)	0.690	(0.33)	0.722	(0.30)	0.712	(0.29)	0.507	(0.36)	0.765	(0.29)
ϕ_{2A}	0.050	(0.02)	0.238	(0.40)	0.164	(0.35)	0.162	(0.35)	0.250	(0.40)	0.150	(0.35)
ϕ_{3A}	-0.124	(0.02)	-0.179	(0.24)	-0.163	(0.22)	-0.157	(0.23)	-0.026	(0.28)	-0.157	(0.22)
μ_B	0.302	(0.10)	0.292	(0.12)	0.297	(0.12)	0.306	(0.12)	0.659	(1.02)	0.296	(0.12)
ϕ_{1B}	0.222	(0.07)	0.233	(0.08)	0.228	(0.08)	0.228	(0.08)	0.255	(0.46)	0.234	(0.07)
ϕ_{2B}	0.131	(0.07)	0.129	(0.08)	0.124	(0.09)	0.131	(0.08)	0.155	(0.45)	0.137	(0.08)
ϕ_{3B}	0.135	(0.07)	0.130	(0.08)	0.142	(0.08)	0.143	(0.08)	0.139	(0.39)	0.137	(0.08)
c	-2.506	(0.08)	-2.462	(0.95)	-2.267	(0.97)	-2.414	(0.97)	0.438	(4.23)	-2.598	(0.76)
α_η or α_θ			0.341				0.157					
Mahalanobis			0.092		0.057		0.241				0.387	
$\rho_{\max}$			0.845				0.895		0.967			
Fn($\times 1000$)			1897		893		1082		35747		113	
#Iter			7		8		6		1		7	

See Table 5 for a description of the entries in the table.

28

The behaviour of the economy in the 'recession' regime A, which occurs in approximately 5% of the observations for most models, can be judged from the relation between the mean of the regime with the largest inverse-root of the AR polynomial. Figure 13 displays μ_A versus $\max_i |\rho_{Ai}^{-1}|$, using the sample from the APMH algorithm. Results for APIS, MH and IS are similar. For the GG sampler the possibility of a non-stationary regime is given very low posterior mass.
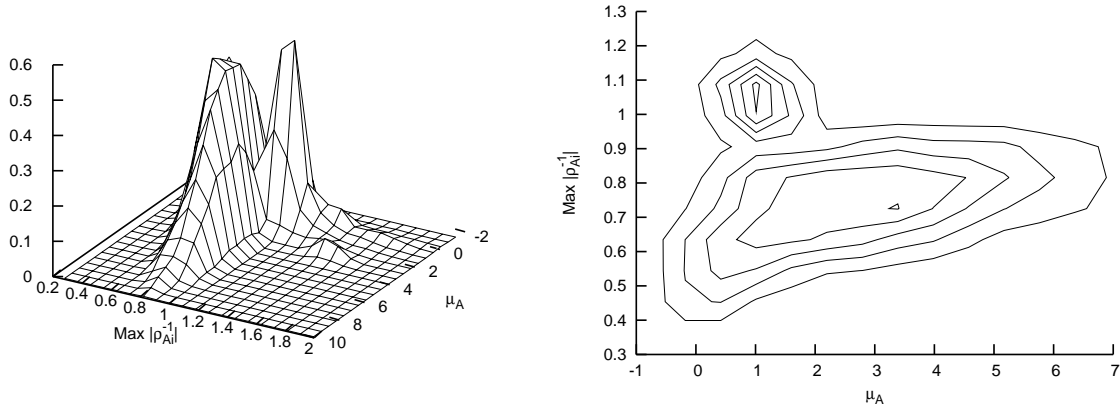

Figure 13: Relation between the mean and the largest inverse root of the AR-process of regime A, for the APMH sample

The acceptance rate which is reached in the APMH algorithm is reasonable: With $\alpha_\eta \approx 0.34$, a sufficient number of directions is accepted to ensure reasonable mixing. The acceptance rate for the MH algorithm is low and the value of the Mahalanobis measure is found to jump up and down, even after apparent convergence of the algorithm. The last value of this Mahalanobis distance is indeed good for both APMH and APIS methods, but not so good for the MH and IS sampling methods.

The number of function evaluations that were used for sampling are interesting as well. For this model, computing a univariate integral is hard due to the non-continuous behaviour of the parameter c (as only the relation of c to the order statistics of the trigger variable s_t is of importance) and of the parameters μ_A, μ_B (taking on different meanings and priors depending on the stability of the AR-polynomial). The effect is that the griddy-Gibbs sampler needs a total of 36 million function evaluations, compared to 1.9 million for the APMH sampler. Though the MH sampler seems to be more efficient in the sense that only one million evaluations were used, its posterior sample shows higher correlation and less convincing convergence judging from the Mahalanobis measure. The APIS algorithm needs less function evaluations; however the posterior sample contains many observations with very low weight.

5   Conclusions

We have extended the Metropolis-Hastings and importance sampling methods by applying a polar transformation to the parameter space of the posterior (or target) density. Sampling does not take place in the m-dimensional parameter space directly, but in an $(m-1)$-dimensional subspace of directions. The last dimension, which corresponds to a distance measure, is sampled exactly from the target density (conditional on the directions), using the

inverse transform method. In this way the shape of the posterior density is taken into account perfectly along the sampled directions. For a given number of draws, this approach requires more functional evaluations of the posterior density than a traditional MH or IS algorithm. The usual type of tradeoff occurs: with a more sophisticated algorithm, one can hope to get 'correct' results with less draws than with a less sophisticated algorithm. It may also happen that a simple method cannot deliver reliable results. It would however be surprising when APS cannot deliver good results while the simpler, less computer intensive methods, can. Using several empirical illustrations in Section 4, this is confirmed. The examples were chosen to illustrate the possibility of using the APS algorithms successfully on a cocktail of econometric models that are of current interest and use. Moreover, a possible use of the APS algorithms is as a preliminary step to explore the posterior distribution and prepare a more sophisticated method.

Let us emphasize that there is no claim that APS algorithms are superior in theory to other kinds of algorithms (such a claim would make no sense). We believe that for any model/data combination, a sufficient research effort will usually allow to find a specific algorithm that performs better than APS or other algorithms. However, this is not necessarily guaranteed, and the specific algorithm may not be better even for a different data set (with the same model).

An interesting extension of this paper would be to embed a polar sampling algorithm in a Gibbs algorithm, where a subset of the parameters can be directly simulated from their conditional distribution, while the remaining parameters cannot. In this framework, special care should be given to start polar sampling with sufficiently good initial guesses of the location and scale of the conditional distribution to be simulated (since at each iteration of the Gibbs sampler, location and scale have in principle to be updated using the last draw of the other parameters). An example where such an algorithm may be of great potential efficiency is in the Bayesian analysis of the linear simultaneous equation model, where the so-called simultaneity parameters induce a very nonelliptical shape of the posterior. Other examples are a cointegration model or some limited dependent variable models.

Appendix 1: Proof of Proposition 1

First, given (15) and (16), the target density $p(x)$ in terms of ρ and η is given by

$$p(\rho, \eta) = p(x(\rho, \eta | \mu, \Sigma)) |J_x(\rho, \eta)| = p(x(\rho, \eta | \mu, \Sigma)) |J_y(\rho)| J_y(\eta) \det(\Sigma^{1/2}), \qquad (35)$$

implying that

$$
\begin{aligned}
p(\eta) &= \int_{-\infty}^{\infty} p(\rho, \eta) \, d\rho = \int_{-\infty}^{\infty} p(x(\rho, \eta | \mu, \Sigma)) |J_y(\rho)| J_y(\eta) \det(\Sigma^{1/2}) \, d\rho \\
&\propto J_y(\eta) \int_{-\infty}^{\infty} p(x(\rho, \eta | \mu, \Sigma)) |J_y(\rho)| \, d\rho
\end{aligned}
\qquad (36)
$$

(the last expression being a kernel of the marginal target density of η). Second, the normal candidate density, denoted by $q(x)$, becomes the following function in terms of ρ and η:

$$
\begin{aligned}
q(\rho, \eta) &= q(x(y(\rho, \eta) | \mu, \Sigma)) |J_x(\rho, \eta)| \\
&\propto \exp\left(-\frac{1}{2} \rho^2\right) |J_y(\rho)| J_y(\eta),
\end{aligned}
\qquad (37)
$$

30

so that[12]

$$q(\eta) \propto J_y(\eta). \tag{38}$$

It follows from (36) and (38) that the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, defined for an independence chain as

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{ \frac{p(\eta_i^*)q(\eta_{i-1})}{p(\eta_{i-1})q(\eta_i^*)}, 1 \right\}, \tag{39}$$

simplifies to the expression in (18). Further, it follows from (35) and (36) that the density of ρ conditional on η is given by (20).

Appendix 2: Sampling setup

Table 9 reports the sampling setup. In total, we run the algorithms at least 4 times with different sample sizes, in order to update the location and scale parameters of the candidate density. After each update, the Mahalanobis distance between the newly sampled location and the previous location is computed as

$$\text{Mahalanobis} = \left(\mu^{(j)} - \mu^{(j-1)} \right)' \left[\Sigma^{(j)} \right]^{-1} \left(\mu^{(j)} - \mu^{(j-1)} \right), \tag{40}$$

where $\mu^{(j)}$ is the estimate of the mean after update j, with $\Sigma^{(j)}$ the corresponding scale estimate.[13] When the Mahalanobis distance is small and no longer changing, this is a sign that no further updating is necessary. However care should be taken that the location and scale estimates are correct: it may happen that the algorithms get trapped in some region, and therefore a low and constant value of the Mahalanobis distance is only a sign of convergence, not a proof. When the Mahalanobis measure drops by more than 50%, we take this as a sign that the location changes considerably, and therefore we update the candidate another time.

Table 9: Sampling setup

Update	1	2	3	≥ 4
Repetitions	1000	1000	5000	10000
Directions	100	200	500	1000
Maximum rejections	3	5	100	200

The second row ('Repetitions') of Table 9 reports the number of parameter vectors θ that are collected, for APMH or APIS, divided over a number of directions η as indicated in the third row of the table (for example, in update 1, for each sampled direction, we sample $1000/100$ distances). In APMH and MH, sampling is only terminated when the number of accepted drawings is equal to the number of repetitions given in Table 9 and the final sample

---

[12]If the candidate is elliptically contoured, ρ and η stay independent and the marginal distribution of η is the same as in (38), but the distribution of ρ is different from what appears in (37).

[13]In the griddy-Gibbs algorithm, no updating is performed, and the Mahalanobis distance is not reported since it is irrelevant.

size is equal to the number of repetitions divided by the acceptance rate. The acceptance rate is called α_η when rejection is on the direction (APMH) and α_θ when it is on the parameter directly (MH) (see the tables of Section 4). The acceptance rate should be preferably not too low (under e.g. 25%), otherwise a strong serial correlation can be expected in the chain of sampled values. This correlation is monitored by the value of ρ_{max}, the maximum value of the first order serial correlation between sampled parameter values. For the methods based on importance sampling, no correlation is introduced into the sample.

In the earlier updates, when the location/scale estimates are not precise yet, APMH and MH tend to get stuck for a long time at the same parameter vector θ or direction η. In order to force the algorithm to move on, a maximum number of consecutive rejections is introduced (see the last line of Table 9), which ensures that the earlier updates give a rapid but rough improvement for the location and scale parameters. In the last update, this procedure is not used.

References

AITKIN, M. (1991): "Posterior Bayes factors," *Journal of the Royal Statistical Society, Series B*, 53, 111–142.

BAUWENS, L., C. S. BOS, AND H. K. VAN DIJK (1999): "Adaptive Polar Sampling, with an Application to a Bayes Measure of Value-at-Risk," Discussion Paper TI 99-082/4, Tinbergen Institute, Erasmus University Rotterdam.

BAUWENS, L., AND M. LUBRANO (1998): "Bayesian inference on GARCH models using the Gibbs sampler," *Econometrics Journal*, 1, C23–C46.

BAUWENS, L., M. LUBRANO, AND J.-F. RICHARD (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

BOS, C. S. (2001): "Time Varying Parameter Models for Inflation and Exchange Rates," Ph.D. thesis, Tinbergen Institute, Erasmus University Rotterdam, TI 256.

BOS, C. S. (2002): "A comparison of marginal likelihood computation methods," *Proceedings of Computational Statistics 2002*.

BOS, C. S., R. J. MAHIEU, AND H. K. VAN DIJK (2000): "Daily Exchange Rate Behaviour and Hedging of Currency Risk," *Journal of Applied Econometrics*, 15(6), 671–696.

BOX, G., AND M. MULLER (1958): "A note on the generation of random normal deviates," *Annals of Mathematical Statistics*, 29, 610–611.

BROWNLEE, K. A. (1965): *Statistical Theory and Methodology in Science and Engineering*. Wiley, New York, 2 edn.

CHAN, K. S., J. D. PETRUCELLI, H. TONG, AND S. W. WOODFORD (1985): "A multiple threshold AR(1) model," *Journal of Applied Probability*, 22, 267–279.

CHIB, S., AND E. GREENBERG (1996): "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12(3), 409–431.

DOORNIK, J. A. (1999): *Object-Oriented Matrix Programming using Ox*. Timberlake Consultants Ltd, London, 3rd edn., See `http://www.nuff.ox.ac.uk/Users/Doornik`.

ENDERS, W., AND C. W. J. GRANGER (1998): "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates," *Journal of Business & Economic Statistics*, 16, 304–311.

FRÜHWIRTH-SCHNATTER, S. (2001): "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching Models," *Journal of the American Statistical Association*, 96(453), 194–209.

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

——— (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–73.

GILKS, W., AND G. ROBERTS (1996): "Strategies for Improving MCMC," in *Markov Chain Monte Carlo in Practice*, ed. by R. S. Gilks, W.R., and D. Spiegelhalter. Chapman and Hall/CRC.

GILKS, W., G. ROBERTS, AND E. GEORGE (1994): "Adaptive direction sampling," *The Statistician*, 43, 179–189.

GIVENS, G. H., AND A. E. RAFTERY (1996): "Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships," *Journal of the American Statistical Association*, 91(433), 132–141.

HAMMERSLEY, J., AND D. HANDSCOMB (1964): *Monte Carlo Methods*. Chapman and Hall, London.

HASTINGS, W. K. (1970): "Monte Carlo Sampling Methods using Markov Chains and their Applications," *Biometrika*, 57, 97–109.

HOBERT, J. P., AND G. CASELLA (1996): "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models," *Journal of the American Statistical Association*, 91(436), 1461–1473.

HOP, J., AND H. VAN DIJK (1992): "SISAM and MIXIN: Two algorithms for the computation of posterior moments and densitites using Monte Carlo integration," *Computer Science in Economics and Management*, 5, 183–220.

JUSTEL, A., AND D. PEÑA (1996): "Gibbs Sampling Will Fail in Outlier Problems with Strong Masking," *Journal of Computational & Graphical Statistics*, 5(2), 176–189.

KASS, R. E., AND A. E. RAFTERY (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

KLEIBERGEN, F. R., AND H. K. VAN DIJK (1994): "On the Shape of the Likelihood/Posterior in Cointegration Models," *Econometric Theory*, 10(3-4), 514–551.

——— (1998): "Bayesian Simultaneous Equations Analysis using Reduced Rank Structures," *Econometric Theory*, 14(6), 701–743.

33

KLOEK, T., AND H. K. VAN DIJK (1978): "Bayesian estimates of equation system parameters: an application of integration by Monte Carlo," *Econometrica*, 46, 1–19.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER (1953): "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1091.

MONAHAN, J., AND A. GENZ (1997): "Spherical-radial integration rules for Bayesian computation," *Journal of the American Statistical Association*, 92, 664–6674.

MONAHAN, J. F. (2001): *Numerical Methods of Statistics*, Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge.

MUIRHEAD, R. (1982): *Aspects of Multivariate Statistical Theory*. Wiley, New York.

OH, M. S., AND J. O. BERGER (1992): "Adaptive Importance Sampling in Monte Carlo Integration," *Journal of Statistical Computation and Simulation*, 41, 143–168.

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

SCHMAISER, B., AND M.-H. CHEN (1991): "General Hit-and-run Monte Carlo Sampling for Evaluating Multidimensional Integrals," Discussion paper, School of Industrial Engineering, Purdue University.

TIERNEY, L. (1994): "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics*, 22, 1701–1762.

VAN DIJK, H. K., AND T. KLOEK (1980): "Further experience in Bayesian analysis using Monte Carlo integration," *Journal of Econometrics*, 14, 307–328.

——— (1984): "Experiments with some alternatives for simple importance sampling in Monte Carlo integration," in *Bayesian Statistics 2*, ed. by J. M. Bernardo, M. Degroot, D. Lindley, and A. F. M. Smith, Amsterdam. North Holland.

VAN DIJK, H. K., T. KLOEK, AND C. G. E. BOENDER (1985): "Posterior Moments Computed by Mixed Integration," *Journal of Econometrics*, 29, 3–18.