THE SIMULATED LIKELIHOOD RATIO METHOD

Monica BILLIO[1]

Alain MONFORT[2]

Christian P. ROBERT[3]

Abstract

A simulation method based on importance sampling and MCMC techniques allows to approximate the likelihood ratio at two different parameter values. Based on this stochastic approximation technique, it is possible to get approximations of the maximum likelihood estimator in the general framework of dynamic latent variable models. Examples of this class of models include factor models, switching regime models, dynamic limited dependent variable models, stochastic volatility models and discretised continuous time models, which all are customarily used on econometrics. We also introduce a rigorous probabilistic setup in order to derive the necessary conditional distributions and several applications show the practical feasibility and the good performance of the proposed approach.

Keywords: Dynamic latent variable model; Instrumental distribution; Markov Chain Monte Carlo method; stochastic approximation.

[1]Università Ca' Foscari, Venice. billio@unive.it

[2]CNAM - CREST, Paris. monfort@ensae.fr

[3]Université Paris Dauphine - CREST, Paris. xian@pi.ceremade.dauphine.fr. Partially supported EU TMR network ERB–FMRX–CT96–0095 on *"Computational and Statistical Methods for the Analysis of Spatial Data"*. The author is thankful to O. Cappé, R. Douc, E. Moulines and N. Shephard for insightful comments.

1 Introduction

The class of *parametric dynamic latent variable models* (also called factor models or hidden variable models or state space models) is becoming increasingly popular, because of the flexibility they offer in the modelling of complex phenomena. These models jointly specify a sequence (y_t) of time dependent variables and a second sequence (y_t^*) of partially unobserved variables in such a way that the joint density $f(y_1, \cdots, y_T, y_1^*, \cdots, y_T^*; \theta)$ is generally much more manageable than the likelihood function $f(y_1, \cdots, y_T; \theta)$, since the later requires a high dimensional integral, whose size is equal to the number of observations multiplied by the dimension of the unobserved variables. The complexity of these models is such that generic latent variable techniques like EM do not apply and they almost invariably call for simulation based methods.

Indeed, simple examples like hidden Markov mixture models, where the unobserved y_t^*'s have a finite support and carry the time dependence, or censored AR models, where the y_t^*'s are distributed from a standard AR(p) model and $y_t = \min\{y_t^*, x_t\}$, can be handled by a variety of approaches, including EM. Now, consider the more involved *stochastic volatility model*, where $(1 \leq t \leq T)$

$$\begin{cases} y_t^* = m_t(y^{*t-1}, y^{t-1}; \theta) + \sigma_t(y^{*t-1}, y^{t-1}; \theta)\varepsilon_t^*, \\ y_t = \exp(0.5\, y_t^*)\varepsilon_t, \end{cases} \tag{1}$$

where y^{*t-1} and y^{t-1} denote $(y_1^*, \ldots, y_{t-1}^*)$ and $(y_1, \ldots, y_{t-1})$, respectively, and where $\{\varepsilon_t^*\}$ and $\{\varepsilon_t\}$ are independent scalar white noises (see Ghysels *et al.* (1996), or Shephard (1996)). Since the y_t^*'s are totally unobserved, the likelihood function is implicit and expressed as a T dimensional integral. Likelihood inference in this model then requires much more specialized tools. The same applies to other models commonly used in econometrics, finance or signal processing, like switching regime dynamic models, factor models or discretized continuous time models.

Simulation based approximation methods can already be found in the literature in the specific case of dynamic latent variable models. A first group of techniques relies on methods which are relatively simple to implement, but which are less efficient than the maximum likelihood approach: see, e.g., the simulated method of moments (McFadden (1989), Duffie and Singleton (1993)), the simulated pseudo-maximum likelihood method (Laroque and Salanié (1993)), the indirect inference method (Gouriéroux, Monfort and Renault (1993)). A second group directly approximates the maximum likelihood estimator via importance sampling (Danielsson and Richard (1993), Billio and Monfort (1998)) or other simulation methods, such as the simulated EM (Shephard (1993, 1994), Diebolt and Ip (1996)). A third approach borrows from Bayesian MCMC algorithms by expressing the maximum likelihood estimator as a limiting Bayes estimate (see, e.g., Robert and Titterington (1998)).

This paper proposes yet another approach, which approximates the likelihood function by MCMC methods in order to compute the maximum of this approximate likelihood. It thus shares the same spirit as Geyer (1994, 1996) and Geyer and Thompson (1992), and is more a stochastic approximation than a stochastic optimization technique. The main idea behind the simulated likelihood ratio method (SLR) is that the (observed) likelihood ratio

$$\frac{f(y_1,\cdots,y_T;\theta)}{f(y_1,\cdots,y_T;\bar\theta)}\ ,$$

where $\bar\theta$ is an arbitrary fixed *reference* value of the parameter, can be expressed as the conditional expectation of a completed likelihood ratio

$$E_{\bar\theta}\left[\frac{f(y_1^*,\ldots,y_{T-k}^*,y_1,\ldots,y_T;\theta)}{f(y_1^*,\ldots,y_{T-k}^*,y_1,\ldots,y_T;\bar\theta)}\middle| y_1,\cdots,y_T\right]\ , \tag{2}$$

where $0\le k\le T$ is an arbitrary integer. A key feature in this representation is that the likelihood ratio appears as the expectation of a completed likelihood ratio with respect to a *single fixed distribution*, namely the conditional distribution of some y_t^*'s given the y_t's for a fixed value $\bar\theta$. (Note that (2) applies for all the usual dynamic latent variables models, including the degenerate cases where the observable variables are deterministic functions of the latent variables.)

An obvious implementation of the SLR method is to evaluate the ratio (2) by Monte Carlo. The simulations from the appropriate distribution, namely the distribution of $y_1^*,\ldots,y_{T-k}^*$ conditional on $y_1,\ldots,y_T$, for a fixed value $\bar\theta$, can be performed using Metropolis-Hastings steps. Note that, given that the instrumental distribution is fixed, the (approximate) objective function is smooth in θ even when observable variables are not differentiable or are discontinuous functions of θ, as in, respectively, switching state space models (Lee, 1997a) or dynamic disequilibrium models (Lee, 1997b), or when the simulation involves a rejection technique. Moreover, since the simulations are done with a fixed distribution, the simulation step is only called once and does not appear at the maximization stage. It can thus be used in a vectorialized procedure computing simultaneously the likelihood ratio at many values of θ; for instance, likelihood surfaces or slices are easily computed.

In theory, the method provides asymptotically efficient estimators of θ and of likelihood ratio test statistics. In practice, the choice of the reference value $\bar\theta$ and of the proposal distribution in the Metropolis-Hastings steps have important bearings on the performance of the approximations. Several proposals are made here, including sequential optimal sampling distributions proposed in Billio and Monfort (1998).

The paper is organised as follows. Section 2 introduces a series of relevant examples, while providing notations for the general setup. (A more rigorous treatment of conditional issues and dominating measures is proposed in Appendix 6.1) Section 3 introduces the general SLR procedure and shows how it applies to various classical models. Section 4 provides several numerical applications. Concluding remarks are given in Section 5

2 Dynamic latent variable models

2.1 Basic assumptions

As already mentioned in the introduction, we consider two separate sets of vectors: the latent variables y_t^*, $t=1,...,T$, are p^*-dimensional vectors of partially or entirely unobservable endogenous variables and the y_t's, $t=1,...,T$, are p-dimensional vectors of observable endogenous variables. (For simplicity's sake, exogenous (explanatory) variables are not mentioned explicitly, although they can be introduced at little cost.) If y^{*t} and y^t denote $(y_1^*, ..., y_t^*)$ and $(y_1, ..., y_t)$, respectively, while $\mathcal{I}_t = \{y^{*t}, y^t\}$, we assume that (y^{*T}, y^T) is absolutely continuous, with joint density $f(y^{*T}, y^T; \theta)$ $\theta \in \Theta \subset \mathrm{R}^k$, with respect to a well-specified product measure (see Appendix 6.1 for details). We denote by $f(y_1^*, y_1; \theta)$ the marginal density of (y_1^*, y_1) and by $f_t(y_t^*, y_t|\mathcal{I}_{t-1}; \theta)$ the density of (y_t^*, y_t) conditional on $\mathcal{I}_{t-1}$. The partial and full conditional densities of y_t^* and y_t, given $\mathcal{I}_{t-1}$, are also denoted by $f_t(y_t^*|\mathcal{I}_{t-1}; \theta)$ and $f_t(y_t|\mathcal{I}_{t-1}; \theta)$, and $f_t(y_t^*|\mathcal{I}_{t-1}, y_t; \theta)$ and $f_t(y_t|\mathcal{I}_{t-1}, y_t^*; \theta)$, respectively. As shown in Appendix 6.1, it is still possible to use Bayes formula in this general setting, namely that

$$f(y^{*T}|y^T; \theta) = \frac{f(y^{*T}, y^T; \theta)}{f(y^T; \theta)}, \tag{3}$$

where $f(y^T; \theta)$ and $f(y^{*T}|y^T; \theta))$ denote the marginal and conditional densities of the vectors y^T and y^{*T}, respectively (see Theorem 6.1 in Appendix 6.1).

We now consider two particular cases of this general setup, which are the most common in practice: the "product" and the "functional" latent models.

2.2 Product latent models

In this case, the joint distribution of (y_t, y_t^*), conditional on the past values (y_j, y_j^*), is dominated by a product measure. A typical occurrence of this setting is when y_t and y_t^* satisfy

$$\begin{cases} y_t^* &= r_t^*(y^{*t-1}, y^{t-1}, \varepsilon_t^*; \theta)\,, \\ y_t &= r_t(y^{*t}, y^{t-1}, \varepsilon_t; \theta)\,, \end{cases} \tag{4}$$

where the ε_t^*'s and ε_t's are independent white noises, with marginal distributions which may depend on θ, and where the ranges of the functions $r_t^*(y^{*t-1}, y^{t-1}, \cdot; \theta)$, and $r_t(y^{*t}, y^{t-1}, \cdot; \theta)$ do not depend on the other arguments $[(y^{*t-1}, y^{t-1}, \theta)$ for r_t^*, and $(y^{*t}, y^{t-1}, \theta)$ for $r_t]$.

We now consider a few examples in this class, some of which have not been satisfactorily treated in the literature so far.

Example 2.1 Factor models. A one factor model is defined by

$$\begin{cases} y_t^* &= r_t^*(y^{*t-1}, y^{t-1}, \varepsilon_t^*; \theta), \\ y_t &= a(\theta) y_t^* + \varepsilon_t, \end{cases} \tag{5}$$

where the ε_t^*'s and ε_t's are iid white noises in R and R^p respectively, which are absolutely continuous wrt their respective Lebesgue measures, λ_1 and λ_p; $a(\theta)$ is a p-dimensional vector of coefficients. If the function $r_t^*(y^{*t-1}, y^{t-1}, \cdot; \theta)$ is one-to-one and differentiable, y_t^* has a probability density function (p.d.f.) (against the Lebesgue measure), $f_t(y_t^*|\mathcal{I}_{t-1}; \theta)$. If $\psi(\varepsilon; \theta)$ is the p.d.f. of ε_t, the conditional density of y_t^* is for instance

$$f_t(y_t^*|\mathcal{I}_{t-1}, y_t; \theta) = \frac{f_t(y_t^*|\mathcal{I}_{t-1}; \theta)\psi[y_t - a(\theta)y_t^*; \theta]}{\int f_t(y_t^*|\mathcal{I}_{t-1}; \theta)\psi[y_t - a(\theta)y_t^*; \theta] d\lambda_1(y_t^*)}.$$

These remarks also apply for the stochastic volatility models (1); we only have to switch from p to 1 and from $\psi[y_t - a(\theta)y_t^*;\theta]$ to $\exp(-0.5\, y_t^*)\psi[y_t \exp(-0.5\, y_t^*);\theta]$.

Example 2.2 Switching regime dynamic models. These models correspond to

$$\begin{cases} y_t^*|y_{t-1}^* & \sim & \pi_{y_{t-1}^* y_t^*}, \\ y_t & = & r_t(y^{*t}, y^{t-1}, \varepsilon_t;\theta), \end{cases} \tag{6}$$

where y_t^* is a K-state Markov chain with transition probabilities π_{ij} $(1 \le i,j \le K)$, which may depend on y^{t-1} and θ, and where ε_t is an absolutely continuous p-dimensional white noise (see Kim (1994), Billio and Monfort (1998), Lee (1997a)). The function $r_t(y^{*t}, y^{t-1}, \cdot;\theta)$ is assumed to be one-to-one and differentiable and y_t given (y^{*t}, y^{t-1}) thus has a p.d.f. $f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta)$. In the particular case K=2, y_t^* is given by

$$y_t^* = \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*),$$

where $\varepsilon_t^* \sim \mathcal{U}_{[0,1]}$, the uniform distribution on [0,1], independent of ε_t. This is to say, y_t^* is a binary variable in $\{0,1\}$. Then

$$\begin{array}{rcl} f_t(y_t^*|\mathcal{I}_{t-1};\theta) & = & \mathbb{1}_{\{0\}}(y_{t-1}^*)[\pi_{00}\mathbb{1}_{\{0\}}(y_t^*) + (1-\pi_{00})\mathbb{1}_{\{1\}}(y_t^*)] \\ & & +\mathbb{1}_{\{1\}}(y_{t-1}^*)[(1-\pi_{11})\mathbb{1}_{\{0\}}(y_t^*) + \pi_{11}\mathbb{1}_{\{1\}}(y_t^*)] \\ f_t(y_t|\mathcal{I}_{t-1};\theta) & = & f_t(y_t|\mathcal{I}_{t-1},0;\theta)[\pi_{00}\mathbb{1}_{\{0\}}(y_{t-1}^*) + (1-\pi_{11})\mathbb{1}_{\{1\}}(y_{t-1}^*)] \\ & & +f_t(y_t|\mathcal{I}_{t-1},1;\theta)[(1-\pi_{00})\mathbb{1}_{\{0\}}(y_{t-1}^*) + \pi_{11}\mathbb{1}_{\{1\}}(y_{t-1}^*)] \\ f_t(y_t^*|\mathcal{I}_{t-1},y_t;\theta) & = & \dfrac{f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta)}{f_t(y_t|\mathcal{I}_{t-1};\theta)}. \end{array} \tag{7}$$

2.3 Functional latent models

In this second type of models, we assume that $y_t = g_t(y_t^*)$, where g_t is a known function; that is, y_t is a deterministic transform of the latent variable, which generally is not one-to-one. So, the conditional distribution of y_t given $(\mathcal{I}_{t-1}, y_t^*)$ is a point mass at $g_t(y_t^*)$. We consider below a few examples in this category.

Example 2.3 Dynamic Probit/Tobit models. We assume that the conditional distribution of y_t^* given $\mathcal{I}_{t-1}$ is absolutely continuous, with conditional p.d.f. $\tilde{f}_t(y_t^*|y^{*t-1};\theta)$. This p.d.f. may be explicit, for instance when y_t^* follows a normal AR(p) model, or it may be given by the Kalman filter if y_t^* is from a normal linear state space model. The observable y_t is then the transform $y_t = \mathbb{1}_{\mathrm{R}^+}(y_t^*)$ in the Probit case and $y_t = \max(y_t^*, 0)$ in the Tobit case (see Manrique and Shephard (1998)). Consider, for instance, the Tobit case. Then, the conditional p.d.f. $f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta)$ is

$$\frac{\mathbb{1}_{\{\mathrm{R}^-\}}(y_t^*)\tilde{f}_t(y_t^*|y^{*t-1};\theta)}{F_{0t}(y^{*t-1};\theta)} \tag{8}$$

where

$$F_{0t}(y^{*t-1};\theta) = \int_{-\infty}^{0} \tilde{f}_t(y|y^{*t-1};\theta)d\lambda_1(y),$$

when $y_t = 0$; otherwise, the conditional distribution of y_t^* is a Dirac mass in y_t.

Example 2.4 Dynamic disequilibrium models. Consider $y_t^* = (y_{1t}^*, y_{2t}^*)$ from an arbitrary bivariate dynamic model. Assume that the conditional distribution of y_t^* given $\mathcal{I}_{t-1}$ has a p.d.f. $\tilde{f}_t(y_t^*|y^{*t-1};\theta)$. The observable y_t is given by $y_t = \min(y_{1t}^*, y_{2t}^*)$ (see Laroque and Salanié (1993), Lee (1997a)). The conditional distribution of y_t^* given $(\mathcal{I}_{t-1}, y_t)$ is then the sum of a Dirac mass at y_t for y_{1t}^* and the distribution with density proportional to $\tilde{f}_t(y_t, y_{2t}^*|y^{*t-1};\theta) \cdot \mathbb{1}_{[y_t,\infty[}(y_{2t}^*)$ for y_{2t}^*, and of a Dirac mass at y_t for y_{2t}^* and the distribution with density proportional to $\tilde{f}_t(y_{1t}^*, y_t|y^{*t-1};\theta) \cdot \mathbb{1}_{[y_t,\infty[}(y_{1t}^*)$ for y_{1t}^*.

Example 2.5 Continuous time processes. Consider the diffusion process $dy_t^* = a(y_t^*;\theta)dt + b(y_t^*;\theta)dW_t$, where W_t is a standard Brownian motion. We assume that this process is observed at regularly spaced intervals. It can then be approximated to an arbitrary degree of precision by the Euler method. If the time unit in the approximation is n times smaller than the lag between two observations, the approximate model is $(t = 1, ..., T = (K-1)n + 1)$.

$$\begin{cases} y_t^* &= y_{t-1}^* + \frac{1}{n}a(y_{t-1}^*;\theta) + \frac{1}{\sqrt{n}}b(y_{t-1}^*;\theta)\varepsilon_t, \\ y_t &= y_t^* \mathbb{1}_{\{n\mathrm{N}\}}(t-1), \end{cases}$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ and $\mathbb{1}_{\{n\mathrm{N}\}}(t)$ is equal to 1 if t is a multiple of n and to 0 otherwise. This model simply states that y_t^* is observed at the K dates 1, $n+1$, $2n+1, \ldots, (K-1)n+1 = T$, that is $y_t =_t^*$ at these dates and $y_t = 0$ otherwise. For this model, the relevant p.d.f.'s are

$$\begin{aligned} f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta) &= \tilde{f}(y_t^*|y_{t-1}^*;\theta) \\ &= \frac{\sqrt{n}}{b(y_{t-1}^*;\theta)}\varphi\left[\frac{\sqrt{n}(y_t^* - y_{t-1}^* - a(y_{t-1}^*;\theta)/n)}{b(y_{t-1}^*;\theta)}\right], \end{aligned}$$

where φ is the standard normal p.d.f., and

$$f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) = \begin{cases} 1 & \text{if } t-1 \in n\mathrm{N}, \\ \tilde{f}(y_t^*|y_{t-1}^*;\theta) & \text{otherwise.} \end{cases} \quad (9)$$

3 The simulated likelihood ratio (SLR) method

3.1 The general principle

The method is based on the following identity (10), which expresses the likelihood ratio as the expectation of a computable ratio. The same device is used by Geyer (1996) in a different setup (see also Geyer and Thompson (1992)); we generalize his result in a dynamic context with potential functional relationships (see §2.3). It also relates to the general principle of importance sampling (Robert and Casella (1999)). (See Appendix 6.2 for a proof of (10).)

Theorem 3.1 *In the setup of Section 2, and for any values θ and $\bar{\theta}$ of the parameter, and any $1 \leq k \leq T$,*

$$\frac{f(y^T;\theta)}{f(y^T;\bar{\theta})} = E_{\bar{\theta}}\left[\frac{f(y^{*T-k}, y^T;\theta)}{f(y^{*T-k}, y^T;\bar{\theta})}\Big| y^T\right], \tag{10}$$

where $f(y^{*T-k}, y^T;\theta)$ is the p.d.f. of (y^{*T-k}, y^T).

As detailed below, this result is particularly useful for the case $k = 0$. However the case $k > 0$ is also interesting when dealing with importance functions of order greater than 1 (see §3.4).

Clearly, the maximization (in θ) of the likelihood function $f(y^T;\theta)$ is equivalent to the maximization (in θ) of the ratio $f(y^T;\theta)/f(y^T;\bar{\theta})$, where $\bar{\theta}$ is a fixed reference value. The complexity of the models under study is such that both $f(y^T;\theta)$ and the ratio are not computable. However, (10) shows that the ratio can be expressed as an expectation and, therefore, that it can be, in principle, approximated by simulations. The Monte Carlo implementation of Theorem 3.1 follows:

---

[1] `Generate` S `simulated paths` $y^{*T-k}(s)$, $s = 1, ..., S$, `from` $f(y^{*T-k}|y^T, \bar{\theta})$.

[2] `Maximize (in` θ`) the average`

$$\frac{1}{S}\sum_{s=1}^{S}\frac{f(y^{*T-k}(s), y^T;\theta)}{f(y^{*T-k}(s), y^T;\bar{\theta})}\,. \tag{11}$$

Note that, in step `[1]`, y^{*T-k} is generated conditionally on y^T and for the reference value $\bar{\theta}$. We call the above algorithm the *Simulated Likelihood Ratio* (SLR) method. A remaining problem is obviously to simulate the $y^{*T-k}(s)$'s from $f(y^{*T-k}|y^T, \bar{\theta})$. We address this problem in the following sections for the specific examples presented in Sections 2.2 and 2.3. But we first point out some distinctive features of the SLR method.

1. For S large enough, the SLR method can be considered as asymptotically efficient, which is not the case for methods like simulated moments, simulated pseudo likelihood or indirect inference.

2. Contrary to simulated EM methods, such as SEM (Celeux and Diebolt (1985), Diebolt and Ip (1996)) and MCEM (Wei and Tanner (1990)), the SLR method only requires one optimization run (as opposed to a sequence of optimizations) and it provides in addition approximations of the likelihood surface and of likelihood ratio test statistics.

3. The simulation step `[1]` operates with a single reference value $\bar{\theta}$ and, therefore, the objective function (11) is smooth in θ, even if the simulation involves rejection methods, if the endogenous variables are not differentiable, or if they are discontinuous functions of the parameter. Moreover, in some settings, when simulation is time consuming, it is interesting to store the simulated paths, instead of generating them anew (with the same seed), when computing (11) for different values

of $\bar{\theta}$. This is the case when convergence of (11) to the observed likelihood ratio is slow, e.g. when $\bar{\theta}$ is far from the maximum likelihood estimator $\hat{\theta}$. As suggested in Geyer (1996), it may then be necessary to repeat the optimization with different values of $\bar{\theta}$, the solution of the maximization program being used as the next $\bar{\theta}$.

4. Since each term of (10) is a ratio of similar quantities for different values of the parameter, numerical problems which can appear in the computation of likelihood functions may be partially avoided. A related issue is variance finiteness: when

$$E_{\bar{\theta}}\left[\left(\frac{f(y^{*T-k}, y^T;\theta)}{f(y^{*T-k}, y^T;\bar{\theta})}\right)^2 \Big| y^T\right] < \infty \tag{12}$$

does not hold, (11) has an infinite variance and may converge to (10) at a very slow rate. However, this problem did not occur in the applications presented below, in the sense that no sudden change in the convergence rate of the likelihood approximation was observed when leaving the parameter region where (12) holds.

3.2 Implementation details

As mentioned above, the method can be implemented by simulating from the conditional distribution of y^{*T-k} given y^T associated with $\bar{\theta}$, i.e. in the distribution defined by the p.d.f. $f(y^{*T-k}|y^T;\bar{\theta})$. Since it is, in general, impossible to simulate directly from this distribution, a natural alternative is to use a Gibbs sampler, for instance to draw successively from the conditional p.d.f.'s $f_t(y^*_t|y^*_{-t}, y^T;\bar{\theta})$, with $y^*_{-t} = (y^*_1, ..., y^*_{t-1}, y^*_{t+1}, ..., y^*_{T-k})$. However, this generation may be difficult to implement.

A more general alternative is to use a Metropolis–Hastings step; that is, at iteration s of the MCMC sampler, to sequentially generate the $z_t(s)$'s ($t = 1, \ldots, T-k$) from the proposal p.d.f.'s $q_t(z_t|z_{t-1}(s), y^{*T-k}(s\text{-}1), y^T)$ and to take the new vector $y^{*T-k}(s)$ as

$$y^{*T-k}(s) = \begin{cases} z^{T-k}(s) & \text{with probability } \rho, \\ y^{*T-k}(s\text{-}1) & \text{with probability } 1-\rho, \end{cases}$$

where ρ is the standard Metropolis–Hastings acceptance probability. Note that, although $f(y^{*T-k}|y^T;\bar{\theta})$ is not available, the ratio involved in ρ can be computed since

$$\frac{f(z^{T-k}(s)|y^T;\bar{\theta})}{f(y^{*T-k}(s\text{-}1)|y^T;\bar{\theta})} = \frac{f(z^{T-k}(s), y^T;\bar{\theta})}{f(y^{*T-k}(s\text{-}1), y^T;\bar{\theta})},$$

and the joint p.d.f. is generally available in closed form.

This global strategy in which the whole vector $z^{T-k}(s)$ is globally accepted or rejected may lead to low acceptance rates, in cases of poor fit between $f(\cdot|y^T;\bar{\theta})$ and the proposal. It then has to be replaced with a multimove strategy in which the sequence is partitioned into blocks, as in Shephard and Pitt (1997). (The theoretical gain resulting from grouping must, however, be supported by a good choice of the proposal.) At iteration s, a block $z^r_p(s) = (z_p(s), ..., z_r(s))$, with $p \le r$, is obtained by first successively drawing in $q_t[z_t|z_p^{t-1}(s), y_p^{*r}(s\text{-}1), y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1)]$, $t = p, ..., r$, with the usual convention that the first set of conditioning variables is empty when $t = p$, and the fourth one is empty when $r = T - k$, and then by taking

$$y_p^{*r}(s) = \begin{cases} z_p^r(s) & \text{with probability } \rho, \\ y_p^{*r}(s\text{-}1) & \text{with probability } 1-\rho, \end{cases}$$

where ρ is equal to

$$\frac{f(z_p^r(s)|y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1), y^T)}{f(y_p^{*r}(s\text{-}1)|y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1), y^T)} \prod_{t=p}^{r} \frac{q_t(y_t^*(s\text{-}1)|y_p^{*t-1}(s\text{-}1), z_p^r(s), y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1))}{q_t(z_t(s)|z_p^{t-1}(s), y_p^{*r}(s\text{-}1), y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1))}, \tag{13}$$

if this quantity is less than 1. Note that again the first ratio is easily computed since it is equal to the ratio of the joint p.d.f.'s

$$\frac{f(y_1^{*p-1}(s), z_p^r(s), y_{r+1}^{*T-k}(s\text{-}1), y^T; \bar\theta)}{f(y_1^{*p-1}(s), y_p^{*r}(s\text{-}1), y_{r+1}^{*T-k}(s\text{-}1), y^T; \bar\theta)}.$$

These joint p.d.f.'s can be factorized using (30) and the terms of this factorization in which no y_t^* appears at dates $t = p, ..., r$ cancel out. The extreme version of this multimove strategy is the "one-move" case in which $r = p$: the first set of conditioning variables in q_t is then empty.

An important issue for the performance of the method pertains to the choice of the proposal q_t. We consider different aspects of this choice in the following paragraphs:

3.3 Proposal of order 1

We take $k = 0$ in (10) and

$$q_t(y_t^*|y^{*t-1}, y^{*T}(s\text{-}1)) = f_t(y_t^*|y^{*t-1}, y^t; \bar\theta), \tag{14}$$

which is an importance sampling p.d.f., and can be called the importance function of order 1 (Billio and Monfort (1998)). Note that q_t does not depend on $y^{*T}(s\text{-}1)$ and therefore leads to an independent Metropolis-Hastings approach (Robert and Casella (1999)). In this case, the global acceptance probability ρ is the minimum of 1 and

$$\begin{aligned} &\frac{f(z^T(s)|y^T; \bar\theta) \prod_{t=1}^{T} f_t(y_t^*(s\text{-}1)|y^{*t-1}(s\text{-}1), y^t; \bar\theta)}{f(y^{*T}(s\text{-}1)|y^T; \bar\theta) \prod_{t=1}^{T} f_t(z_t(s)|z^{t-1}(s), y^t; \bar\theta)} \\ &= \frac{\prod_{t=1}^{T} f_t(y_t|z^{t-1}(s), y^{t-1}; \bar\theta)}{\prod_{t=1}^{T} f_t(y_t|y^{*t-1}(s\text{-}1), y^{t-1}; \bar\theta)}. \end{aligned} \tag{15}$$

In some settings, both the simulation from (14) and the computation of (15) are manageable: this is the case for some factor and switching regime dynamic models, and the dynamic Probit/Tobit, dynamic disequilibrium, and continuous time models.

Example 3.2 Factor ARCH models. Consider, for instance, a normal factor ARCH model defined by

$$\begin{cases} y_t^* = (\alpha + \beta y_{t-1}^{*2})^{1/2}\varepsilon_t^*, \\ y_t = a y_t^* + \varepsilon_t, \end{cases} \tag{16}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,\Sigma)$ independently and (a,α) satisfies some identifying condition (for instance, the first component of a is 1 or $\alpha = 1$). We have, with $\theta = (\alpha,\beta,a,\Sigma)$ and $\bar\theta = (\bar\alpha,\bar\beta,\bar a,\bar\Sigma)$,

$$\begin{aligned} f_t(y_t^*|y^{*t-1},y^t;\bar\theta) &\propto \exp\Big\{y_t^*\bar a'\bar\Sigma^{-1}y_t \\ &\quad -\frac{1}{2}y_t^{*2}\left[\frac{1}{\bar\alpha+\bar\beta y_{t-1}^{*2}} + \bar a'\bar\Sigma^{-1}\bar a\right]\Big\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_t^2}[y_t^* - m_t]^2\right\}, \end{aligned}$$

with

$$\begin{aligned} \sigma_t^2 &= \left[\frac{1}{\bar\alpha+\bar\beta y_{t-1}^{*2}} + \bar a'\bar\Sigma^{-1}\bar a\right]^{-1}, \\ m_t &= \bar a'\bar\Sigma^{-1}y_t\sigma_t^2. \end{aligned}$$

Therefore,

$$f_t(y_t^*|y^{*t-1},y^t;\bar\theta) = \frac{1}{\sigma_t}\varphi\left(\frac{y_t^*-m_t}{\sigma_t}\right),$$

Similarly the p.d.f. $f_t(y_t|y^{*t-1},y^{t-1};\bar\theta)$ involved in ρ is a normal density with mean zero and covariance matrix $(\bar\alpha+\bar\beta y_{t-1}^{*2})\bar a\bar a' + \bar\Sigma$. Therefore the previous approach applies.

Example 3.3 Switching regime dynamic models. In these models, the p.d.f.'s, $f_t(y_t^*|y^{*t-1},y^t;\bar\theta)$ are easily computed, provided that the p.d.f.'s $f_t(y_t|y^{*t},y^{t-1};\bar\theta)$ are available (see (7)). It is the case in the general framework of switching state space models (Billio and Monfort (1998)), where these p.d.f.'s are obtained through the partial Kalman filter. It is also the case in a switching GARCH process like

$$\begin{aligned} y_t^* &= 1\!\!1_{\{0\}}(y_{t-1}^*)1\!\!1_{[\pi_{00},1]}(\varepsilon_t^*) + 1\!\!1_{\{1\}}(y_{t-1}^*)1\!\!1_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \sigma_t\varepsilon_t, \\ \sigma_t^2 &= \alpha + \beta y_t^* + \gamma y_{t-1}^2 + \delta\sigma_{t-1}^2, \end{aligned} \tag{17}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{U}_{[0,1]}$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$ independently, and $\theta = (\alpha,\beta,\gamma,\delta,\pi_{00},\pi_{11})$. Here $f_t(y_t|y^{*t},\ y^{t-1};\bar\theta)$ is the normal p.d.f. that would be obtained for the GARCH process if the y_t^*'s were deterministic, i.e. the normal p.d.f. with zero mean and variance $\sigma_t^2(y^{*t},y^{t-1},\alpha,\beta,\gamma,\delta)$ obtained by solving recursively the difference equation given in the model definition. Since we are in a discrete case, simulations of the y_t^*'s are straightforward. Similarly, the p.d.f.'s $f_t(y_t|y^{*t-1},y^{t-1};\bar\theta)$ are available (see (7)) and, therefore, the acceptance probability ρ, given in (15), is easily derived.

Example 3.4 Dynamic Tobit models. In the dynamic Tobit case, the conditional distribution of y_t^* given (y^{*t-1}, y^t) is given by (8). The simulation in this distribution is easily done, since $z_t(s) = y_t$ when $y_t > 0$, and

$$z_t(s) \sim \frac{f_t(z_t|z^{t-1}(s);\bar{\theta})}{F_{0t}(z^{t-1}(s);\bar{\theta})} 1\!\!1_{\mathrm{R}^-}(z_t)$$

if $y_t = 0$. The second case is thus simulated from a truncated distribution. We know that this can be done smoothly in the normal case but, as noted in Section 3, the SLR method does not require smoothness and, in non-normal cases, rejection methods can be used. (See Geweke (1991) and Philippe (1997) for optimal generation algorithms in the truncated normal and Gamma cases, respectively.) The p.d.f.'s $f_t(y_t|y^{*t-1}, y^{t-1};\bar{\theta})$ are simply (see (8))

$$\begin{cases} F_{0t}(y^{*t-1}(s);\bar{\theta}) & \text{if } y_t = 0, \\ \tilde{f}_t(y_t|y^{*t-1}(s);\bar{\theta}) & \text{if } y_t > 0. \end{cases}$$

Therefore ρ is easily computed.

Example 3.5 Dynamic disequilibrium models. In the dynamic disequilibrium model, the conditional distribution of y_t^* given (y^{*t-1}, y^t) is given in Example 2.4. The simulation from this distribution is straightforward:

---

[1] Draw the regimes:

1: with probability

$$\pi = \frac{\int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|z^{t-1}(s);\bar{\theta}) d\lambda_1(y_2^*)}{\int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|z^{t-1}(s);\bar{\theta}) d\lambda_1(y_2^*) + \int_{[y_t,\infty[} \tilde{f}_t(y_1^*, y_t|z^{t-1}(s);\bar{\theta}) d\lambda_1(y_1^*)}$$

2: with probability $1-\pi$,

[2] Draw $z_t(s)$ as

regime 1: $z_{1t}(s) = y_t$ and

$$\tilde{z}_{2t}(s) \sim \frac{\tilde{f}_t(y_t, z_{2t}|z^{t-1}(s);\bar{\theta})}{\int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|z^{t-1}(s);\bar{\theta}) d\lambda_1(y_2^*)} \cdot \lambda_{[y_t,\infty[}$$

regime 2: $z_{2t}(s) = y_t$ and

$$\tilde{z}_{1t}(s) \sim \frac{\tilde{f}_t(z_{1t}, y_t|z^{t-1}(s);\bar{\theta})}{\int_{[y_t,\infty[} \tilde{f}_t(y_1^*, y_t|z^{t-1}(s);\bar{\theta}) d\lambda_1(y_1^*)} \cdot \lambda_{[y_t,\infty[}$$

In the normal case, the previous generations correspond to truncated normal distributions. The p.d.f.'s $f_t(y_t|y^{*t-1}, y^{t-1}; \bar{\theta})$ are given in Example 2.4 and, therefore, the acceptance probability ρ is easily computed.

Example 3.6 Continuous time models. For a continuous time model as in Example 2.5, the conditional distribution of y_t^* given (y^{*t-1}, y^t) is the Dirac mass at y_t if t-1$\in n\mathbb{N}$, and the distribution with density $\tilde{f}(y_t^*|y_{t-1}^*; \bar{\theta})$ otherwise (see (9)). The acceptance probability is then

$$\min \left\{ 1, \prod_{k=2}^{K} \frac{\tilde{f}(y_{(k-1)n+1}|z_{(k-1)n}(s); \bar{\theta})}{\tilde{f}(y_{(k-1)n+1}|y^*_{(k-1)n}(s\text{-}1); \bar{\theta})} \right\}.$$

In this case it is natural to consider a multimove strategy in which the blocks are of length n. More precisely we consider a block corresponding to dates $(k$-1$)n$+1,...,kn. At iteration s the proposed sequence is generated successively from $\tilde{f}(z_{(k-1)n+1}|y^*_{(k-1)n}(s); \bar{\theta})$ and from $\tilde{f}(z_t|z_{t-1}(s); \bar{\theta})$ $(t = (k-1)n + 2, ..., kn)$. The acceptance probability is

$$\rho = \min \left\{ 1, \frac{A[z_{(k-1)n+1}^{kn}(s)]}{A[y_{(k-1)n+1}^{*kn}(s\text{-}1)]} \right\},$$

with

$$A[z_{(k-1)n+1}^{kn}(s)] = \tilde{f}(y_{kn+1}|z_{kn}(s); \bar{\theta}),$$

that is,

$$\frac{\tilde{f}(z_{(k-1)n+1}(s)|y^*_{(k-1)n}(s); \bar{\theta})}{\tilde{f}(z_{(k-1)n+1}(s)|y^*_{(k-1)n}(s); \bar{\theta})} \times \prod_{t=(k-1)n+2}^{kn} \frac{\tilde{f}(z_t(s)|z_{t-1}(s); \bar{\theta})\tilde{f}(y^*_{kn+1}(s\text{-}1)|z_{kn}(s); \bar{\theta})}{\tilde{f}(z_t(s)|z_{t-1}(s); \bar{\theta})}.$$

Therefore

$$\rho = \min \left\{ 1, \frac{\tilde{f}(y_{kn+1}|z_{kn}(s); \bar{\theta})}{\tilde{f}(y_{kn+1}|y^*_{kn}(s\text{-}1); \bar{\theta})} \right\}.$$

3.4 Proposal of order 2

We also consider another case of interest, when $k = 1$ in (10) and

$$q_t(y_t^*|y^{*t-1}, y^{*T}(s\text{-}1)) = f_t(y_t^*|y^{*t-1}, y^{t+1}; \bar{\theta}),$$

where

$$f_t(y_t^*|y^{*t-1}, y^{t+1}; \bar{\theta}) \propto f_{t+1}(y_{t+1}|y^{*t}, y^t; \bar{\theta}) f_t(y_t^*|y^{*t-1}, y^t; \bar{\theta})$$

(see Appendix 6.3). We can use q_t in an independent Metropolis–Hastings approach to draw $z_t(s)$, $t = 1, ..., T-1$. The acceptance ratio is the minimum of 1 and

$$\rho = \min\left\{1, \frac{f(z^{T-1}(s)|y^T;\bar{\theta}) \prod_{t=1}^{T-1} f_t(y_t^*(s\text{-}1)|y^{*t-1}(s\text{-}1), y^{t+1};\bar{\theta})}{f(y^{*T-1}(s\text{-}1)|y^T;\bar{\theta}) \prod_{t=1}^{T-1} f_t(z_t(s)|z^{t-1}(s), y^{t+1};\bar{\theta})}\right\}. \tag{18}$$

But, since

$$\begin{aligned} f(y^{*T-1}|y^T;\bar{\theta})f(y^T;\bar{\theta}) &= f(y^{*T-1}, y^T;\bar{\theta}) \\ &= f_1(y_1;\bar{\theta}) \prod_{t=1}^{T-1} f_t(y_t^*, y_{t+1}|y^{*t-1}, y^t;\bar{\theta}) \\ &= f_1(y_1;\bar{\theta}) \prod_{t=1}^{T-1} f_t(y_t^*|y^{*t-1}, y^{t+1};\bar{\theta}) f_{t+1}(y_{t+1}|y^{*t-1}, y^t;\bar{\theta}), \end{aligned}$$

we get

$$\rho = \min\left\{1, \frac{\prod_{t=1}^{T-1} f_{t+1}(y_{t+1}|z^{t-1}(s), y^t;\bar{\theta})}{\prod_{t=1}^{T-1} f_{t+1}(y_{t+1}|y^{*t-1}(s\text{-}1), y^t;\bar{\theta})}\right\}.$$

This method applies if it is easy to simulate from $f_t(y_t^*|y^{*t-1}, y^{t+1};\bar{\theta})$ and if $f_{t+1}(y_{t+1}|y^{*t-1}, y^t;\bar{\theta})$ is easily computable; this is generally the case when y_t^* takes a finite number of values (see Billio and Monfort (1998) for the switching state space case).

3.5 Proposal with future values of y_t^*

We take $k = 0$ in (10) and

$$q_t(y_t^*|y^{*t-1}, y^{*T}(s\text{-}1)) \propto \begin{cases} f_t(y_t^*|y^{*t-1}, y^{t-1};\bar{\theta}) f_{t+1}(y_{t+1}^*(s\text{-}1)|y^{*t}, y^t;\bar{\theta}), & \\ \qquad \times f_t(y_t|y^{*t}, y^{t-1};\bar{\theta}), & \text{if } t < T, \\ f_t(y_t^*|y^{*t-1}, y^{t-1};\bar{\theta}) f_t(y_t|y^{*t}, y^{t-1};\bar{\theta}), & \text{if } t = T. \end{cases}$$

More precisely, we get

$$\begin{aligned} q_t(z_t|z^{t-1}(s), y^{*T}(s\text{-}1)) \propto\ & f_t(z_t|z^{t-1}(s), y^{t-1};\bar{\theta}) f_{t+1}(y_{t+1}^*(s\text{-}1)|z_t, z^{t-1}(s), y^t;\bar{\theta}) \\ & \times f_t(y_t|z_t, z^{t-1}(s), y^{t-1};\bar{\theta}), \end{aligned}$$

with the convention that the second term is equal to 1 for $t = T$. This p.d.f. will be known exactly if it is possible to compute the normalizing constant. This is easily done, for instance, if y_t^* can take a finite number of values, which is the case in the switching regime models. If y_t^* can take the values 0 and 1, we have

$$\begin{aligned} & q_t(z_t|z^{t-1}(s), y^{*T}(s\text{-}1)) \\ & = \frac{f_t(z_t|z^{t-1}(s), y^{t-1};\bar{\theta}) f_{t+1}(y_{t+1}^*(s\text{-}1)|z_t, z^{t-1}(s), y^t;\bar{\theta}) f_t(y_t|z_t, z^{t-1}(s), y^{t-1};\bar{\theta})}{\sum_{z=0}^{1} f_t(z|z^{t-1}(s), y^{t-1};\bar{\theta}) f_{t+1}(y_{t+1}^*(s\text{-}1)|z, z^{t-1}(s), y^t;\bar{\theta}) f_t(y_t|z, z^{t-1}(s), y^{t-1};\bar{\theta})}. \end{aligned}$$

Since

$$f(z^T(s)|y^T) = \prod_{t=1}^{T} f_t(z_t(s)|z^{t-1}(s), y^{t-1}; \bar{\theta}) f_t(y_t|z_t(s), z^{t-1}(s), y^{t-1}; \bar{\theta}) / f(y^T) \,,$$

$\rho = \min\{1, \rho^*\}$, with

$$\rho^* = \frac{\prod_{t=1}^{T-1} f_{t+1}(z_{t+1}(s)|y^{*t}(s\text{-}1), y^t; \bar{\theta})}{\prod_{t=1}^{T-1} f_{t+1}(y^*_{t+1}(s\text{-}1)|z^{*t}(s), y^t; \bar{\theta})} \times \frac{\prod_{t=1}^{T} \sum_{z=0}^{1} f_t(z|z^{t-1}(s), y^{t-1}; \bar{\theta}) f_{t+1}(y^*_{t+1}(s\text{-}1)|z, z^{t-1}(s), y^t; \bar{\theta}) f_t(y_t|z, z^{t-1}(s), y^{t-1}; \bar{\theta})}{\prod_{t=1}^{T} \sum_{z=0}^{1} f_t(z|y^{*t-1}(s\text{-}1), y^{t-1}; \bar{\theta}) f_{t+1}(z_{t+1}(s)|z, y^{*t-1}(s\text{-}1), y^t; \bar{\theta}) f_t(y_t|z, y^{*t-1}(s\text{-}1), y^{t-1}; \bar{\theta})},$$

(see Billio, Monfort and Robert (1996) for an application of this Metropolis–Hastings approach in the Bayesian framework). Note that this approach is no longer an independent approach. It can be used for the switching state space and GARCH models considered in Examples 3.2 and 3.3.

3.6 Use of approximations

The proposal can also be based on some approximations of $f_t(y_t^*|y_{-t}^*, y^T; \bar{\theta})$.

Example 3.7 Stochastic volatility models. The most classical examples of (1) are the basic model

$$\begin{cases} y_t^* = a + by_{t-1}^* + c\varepsilon_t^*, \\ y_t = \exp(0.5\, y_t^*)\varepsilon_t, \end{cases} \tag{19}$$

where ε_t^* and ε_t are normal independent white noises, and the leverage effect model

$$\begin{cases} y_t^* = a + by_{t-1}^* + c\tilde{\varepsilon}_t, \\ y_t = \exp(0.5\, y_t^*)\varepsilon_t, \end{cases} \tag{20}$$

where $(\tilde{\varepsilon}_t, \varepsilon_{t-1})$ is a normal white noise such that

$$(\tilde{\varepsilon}_t, \varepsilon_{t-1}) \sim \mathcal{N}\left[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

with $\rho < 0$. The model (20) can be expressed as (19) by considering the regression of $\tilde{\varepsilon}_t$ on ε_{t-1} and, for simplicity's sake, we only consider (19). The simulation of the y_t^*'s can proceed according to several approximations of the p.d.f. $f(y_t^*|y_{t-1}^*, y_{t+1}^*, y^t, \theta)$, proportional to

$$\exp\Big\{-(y_t^* - a - by_{t-1}^*)^2/2c^2 \\ -(y_{t+1}^* - a - by_t^*)^2/2c^2 - y_t^*/2 - y_t^2 e^{-y_t^*}/2\Big\}\,. \tag{21}$$

First, the part $(y_t^* + y_t^2 e^{-y_t^*})/2$ in the exponential can be linearized as in Shephard and Pitt (1997). That is, it can be approximated by $(y_t^* - \log(y_t^2))^2/4$ (up to an additive constant) by a second order series expansion around $\log(y_t)^2$. The proposal distribution is then a normal distribution with mean

$$\frac{[a + by_{t-1}^* + b(y_{t+1}^* - a)]c^{-2} + \log(y_t^2)/2}{(1+b^2)c^{-2} + 1/2}$$

and variance $[(1+b^2)c^{-2} + 1/2]^{-1}$, with the appropriate Metropolis-Hastings acceptance probability.

A second solution follows from Jacquier, Polson and Rossi (1994) and it is based on a Gamma approximation for $\omega_t = \exp(-y_t^*)$. In fact, the distribution on ω_t corresponding to (21) is proportional to

$$\omega_t^{-1/2} \exp(-y_t^2 \omega_t/2)\, e^{-(-\log(\omega_t))^2(1+b^2)/2c^2} \omega_t^{-[a+b(y_{t-1}^* + y_{t+1}^* - a)]c^{-2}}.$$

We can then approximate the second part of this distribution by a Gamma distribution $\mathcal{G}(\alpha, \beta)$ on ω_t, which means $\omega_t \sim \pi(\omega_t) \propto \omega_t^\alpha \exp(-\beta\omega_t)$, by either matching the two first moments, as in Jacquier *et al.* (1994), or by a Taylor expansion of $\alpha y_t^* + \beta e^{-y_t^*} = \alpha \log(\omega_t) - \beta\omega_t$.

4 Applications

4.1 Switching regime dynamic models

Switching regime dynamic models have generated considerable attention in the econometric literature. Hamilton (1988), (1989) and many others have used autoregressions with Markov regime switching to modelize various economic phenomena. However, when the autoregression is replaced by an ARMA, Hamilton's approach no longer works. Such extensions have been considered by Lam (1990), who deals with a very special case for which it is possible to apply the Hamilton's filter, Kim (1994) who suggests a recursive approximation of the likelihood, while Lee (1997a) and Billio and Monfort (1998) propose a simulated maximum likelihood approach.

We first consider the simplest case where the global Markovian property is lost and where Hamilton's algorithm does not apply, which is the switching MA(1) model. More precisely,

$$\begin{aligned} y_t^* &= \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \mu_0 + (\mu_1 - \mu_0)y_t^* + \sigma\varepsilon_t + \alpha\sigma\varepsilon_{t-1}, \end{aligned} \tag{22}$$

where $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$, $\varepsilon_t^* \overset{iid}{\sim} \mathcal{U}[0,1]$ independently, $\varepsilon_0 = 0$ and y_1^* is generated from the invariant distribution of the Markov chain.

We use a Metropolis–Hastings algorithm, with an importance sampling p.d.f. of order 1, which also takes into account future values of the latent variable y_t^* (see §3.5 and Billio *et al.* (1998)).

`Set` $\eta_0^{(s)} = 0$.
`Generate` z_0 `from the invariant distribution associated with` $(\bar\pi_{00}, \bar\pi_{11})$.
`For` $t = 1, \ldots, T-1$,

1. `Generate` $z_t \in \{0,1\}$ `from`

$$p(z_t) \propto \exp\{-(y_t - \bar\mu_{z_t} - \bar\alpha\eta_{t-1}^{(s)})^2/2\bar\sigma^2\}\bar\sigma^{-1}\bar\pi_{z_{t-1}z_t}\bar\pi_{z_t y^*_{t+1}}$$

2. `Actualize` $\eta_t^{(s)} = y_t - \bar\mu_{z_t} - \bar\alpha\eta_{t-1}^{(s)}$

`Generate` z_T `from`

$$\exp\{-(y_T - \bar\mu_{z_T} - \bar\alpha\eta_{T-1}^{(s)})^2/2\bar\sigma^2\}\bar\sigma^{-1}\bar\pi_{z_{T-1}z_T} \ .$$

The resulting simulated regimes are proposals for a new allocation of the observations and they are jointly accepted as the new value of y^{*T} with probability ρ equal to the minimum of 1 and

$$\prod_{t\geq 1} \frac{\bar\pi_{y^*_t z_{t+1}}}{\bar\pi_{z_t y^*_{t+1}}} \times \frac{\sum_{i=0}^{1} \exp\{-(y_t - \bar\mu_i - \bar\alpha\eta_{t-1}^{(s)})^2/2\bar\sigma^2\}\bar\pi_{z_{t-1}i}\bar\pi_{i y^*_{t+1}}}{\sum_{i=0}^{1} \exp\{-(y_t - \bar\mu_i - \bar\alpha\eta_{t-1}^{(s-1)})^2/2\bar\sigma^2\}\bar\pi_{y^*_{t-1}i}\bar\pi_{i z_{t+1}}},$$

where $\eta_t^{(s-1)}$ is iteratively defined by $\eta_0^{(s-1)} = 0$ and

$$\eta_t^{(s-1)} = y_t - \bar\mu_{y^*_t} - \bar\alpha\eta_{t-1}^{(s-1)} \ .$$

In order to evaluate the performance of the SLR method through a Monte Carlo experiment, we select two increasingly mixed-up models . We consider 50 simulated samples of size $T = 200$ and 2000 iterations of the completion step, of which we discard the first 1000. Some simulated samples are presented in Figure 1. The different parameters of these models are described in Tables 1 and 2 (first column), along with the reference values ($\bar\theta$, second column) and the estimates obtained by numerical maximization of the likelihood ratio (third column), their Monte Carlo standard deviation and their mean square error (fourth and fifth columns, respectively). Since we are in a simulation setup, we can also compare the estimated regimes versus the true ones (given the parameter estimates, 1500 iterations of the completion step are performed and the average over the last 500 is calculated for every time t: if this average is over 0.5, we consider $y^*_t = 1$ and take $y^*_t = 0$ otherwise). In both cases, the correct assignment rate (adequation rate) is around 90%.

We also consider a MA(2) model with switching mean,

$$\begin{aligned} y^*_t &= \mathbb{1}_{\{0\}}(y^*_{t-1})\mathbb{1}_{[\pi_{00},1]}(\varepsilon^*_t) + \mathbb{1}_{\{1\}}(y^*_{t-1})\mathbb{1}_{[0,\pi_{11}]}(\varepsilon^*_t), \\ y_t &= \mu_0 + (\mu_1 - \mu_0)y^*_t + \sigma\varepsilon_t + \alpha\sigma\varepsilon_{t-1} + \varphi\sigma\varepsilon_{t-2}, \end{aligned} \quad (23)$$

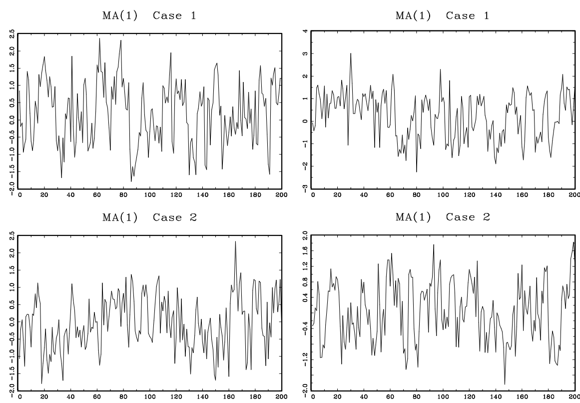

Figure 1: Simulated samples from the MA(1) model (22), cases 1 and 2.

	True	Starting value	SLR estimate	Standard deviation	Mean square error
μ_0	-0.72	-1.2	-0.73775	0.069128	0.005094
μ_1	0.94	1.2	0.90135	0.072316	0.006723
α	0.32	0	0.250742	0.078074	0.010892
σ	0.5	0.7	0.497563	0.029288	0.000864
π_{00}	0.8	0.6	0.777373	0.043827	0.002433
π_{11}	0.8	0.6	0.785902	0.047531	0.002458

Table 1: MA(1) model (22) - case 1 - MC replications 50 - acceptance rate 0.59 - adequation rate 0.94.

	True	Starting Value	SLR Estimate	Standard Deviation	Mean Square Error
μ_0	-0.52	-0.8	-0.5189379	0.08007504	0.00641314
μ_1	0.64	0.8	0.57466772	0.07325565	0.0096347
α	0.32	0	0.25310786	0.08338503	0.01142762
σ	0.5	0.7	0.51042218	0.0328476	0.00118759
π_{00}	0.8	0.6	0.73542635	0.05132958	0.00680448
π_{11}	0.8	0.6	0.75623976	0.05240122	0.00466085

Table 2: MA(1) model (22) - case 2 - MC replications 50 - acceptance rate 0.38 - adequation rate 0.88.

	True	Starting Value	SLR Estimate	Standard Deviation	Mean Square Error
μ_0	-0.72	-1.2	-0.74947	0.053353	0.003715
μ_1	0.94	1.2	0.91122	0.060264	0.00446
α	0.32	0	0.145487	0.101206	0.040698
φ	-0.52	0	-0.26113	0.073926	0.072478
σ	0.5	0.7	0.544187	0.029492	0.002822
π_{00}	0.8	0.6	0.771992	0.037666	0.002203
π_{11}	0.8	0.6	0.763287	0.047319	0.003587

Table 3: MA(2) model (23) - MC replications 50 - acceptance rate 0.67 - adequation rate 0.93.

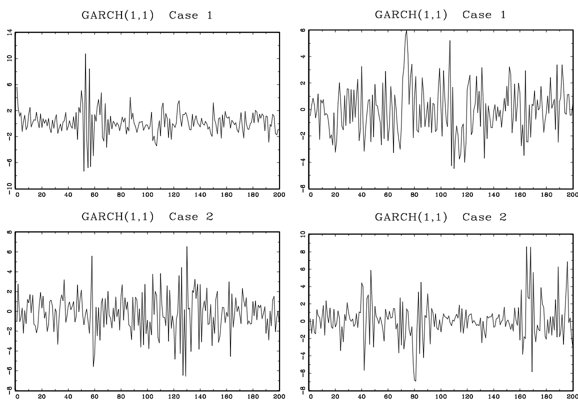

Figure 2: Simulated samples from the GARCH(1,1) model (24), cases 1 and 2.

and we use the same Metropolis-Hastings algorithm. The setup of the Monte Carlo simulation is the same as above, with $T = 200$. The results are shown in Table 3.

As in the MA(1) case, except for α and φ, all the parameters are well estimated and the adequation rate is again over 90%. In all cases we found that the reference value $\bar{\theta}$ has small influence, except for α and φ. In any case, it is possible to select a good reference value $\bar{\theta}$ such as the Bayes estimates obtained by Billio *et al.* (1998).

Finally we consider a GARCH(1,1) model with switching constant in the variance,

$$\begin{aligned} y_t^* &= \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha + \beta y_t^* + \gamma y_{t-1}^2 + \delta \sigma_{t-1}^2, \end{aligned} \tag{24}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{U}_{[0,1]}$ and $\varepsilon_t \overset{iid}{\sim} \mathcal{N}(0,1)$ independently and y_1^* is generated from the invariant distribution of the Markov chain. In this case we consider a Metropolis-Hastings algorithm, with an importance sampling p.d.f. of order 1, as described in Section 3.3, in order to draw from the conditional distribution of y^{*T} given y^T. As for the MA models, we select two increasingly mixed-up models to study the behaviour of the SLR method. We consider again simulated samples of size T=200 and perform 50 replications, with 2000 iterations of the completion step, of which we use only the last 1000 ones to estimate

	True	Starting value	SLR estimate	Standard deviation	Mean square error
α	0.1	0.2	0.22443462	0.18812571	0.05087526
β	0.5	0.2	0.59256369	0.3839882	0.15601498
γ	0.3	0.4	0.29659416	0.1119298	0.01253988
δ	0.6	0.5	0.55842042	0.12173233	0.01654762
π_{00}	0.8	0.7	0.75101444	0.07040026	0.00735578
π_{11}	0.85	0.7	0.77957762	0.08059188	0.01145436

Table 4: GARCH(1,1) model (24) - case 1 - MC replications 50 - only two regime cases: 45 - acceptance rate 0.55 - adequation rate 0.58.

	True	Starting value	SLR estimate	Standard deviation	Mean square error
α	0.1	0.2	0.244662	0.218375	0.068615
β	0.8	0.5	0.84713	0.539081	0.292829
γ	0.3	0.4	0.29776	0.10745	0.01155
δ	0.6	0.5	0.565509	0.133849	0.019105
π_{00}	0.8	0.7	0.737477	0.076355	0.009739
π_{11}	0.85	0.7	0.766831	0.086	0.014313

Table 5: GARCH(1,1) model (24) - case 2 - MC replications 50 - only two regime cases: 48 - acceptance rate 0.24 - adequation rate 0.6.

the likelihood ratio.

The structure of the model is quite involved (as can be seen from Figure 2) and in some cases it is not possible to distinguish the two different regimes (5 times out of 50 in case 1 and only 2 times out of 50 in case 2), therefore we present the estimation results only in the two regime cases in Tables 4 and 5. The two examples show a quite satisfactory result since the SLR estimates are close to the true parameters and the adequation rate is around 60%.

As far as the reference value is concerned, it is important to consider reasonable values of the parameters. It is possible to obtain a valuable reference value by considering the estimates of the parameter obtained with the approximation of the likelihood suggested by Kim (1994) and Dueker (1997).

4.2 Dynamic Tobit models

We consider the following dynamic Tobit model,

$$\begin{cases} y_t^* = \alpha + \beta y_{t-1}^* + \sigma \varepsilon_t^*, \\ y_t = \max(y_t^*, 0), \end{cases} \tag{25}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{N}(0,1)$. In this setting, Manrique and Shephard (1998) devise an MCMC simulated EM algorithm and derive the Bayes estimates. In the simulations we use a Metropolis-Hastings algorithm, with an importance function of order 1, as described in Example 3.4. We consider 50 simulated samples of size T=200, 2000 iterations of the Metropolis-Hastings algorithm, of which we discard the first 1000 ones in order to

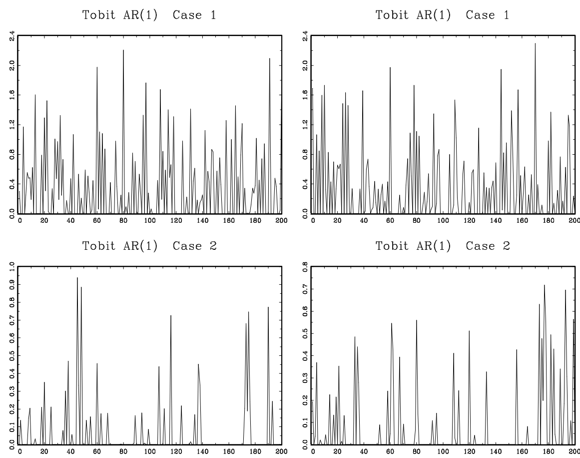

Figure 3: Simulated samples from the Tobit AR(1) model (25), cases 1 and 2.

	True	Starting Value	SLR Estimate	Standard Deviation	Mean Square Error
α	0	0.5	0.0455	0.070522	0.007044
β	-0.4	0.3	-0.32567	0.046162	0.007656
σ	0.8	1	0.780928	0.060573	0.004033

Table 6: Tobit AR(1) model (25) - case 1 - MC replications 50 - acceptance rate 0.51 - zero percentage 0.4978.

estimate the likelihood ratio. We select two different parameter sets, the second one with an higher zero percentage. Some simulated samples are presented in Figure 4.1.

In both cases the results are quite interesting (see Tables 6 and 7): the parameter estimates are close to the true values and there is no dependence on the reference value.

We also analyze a more complicated model, where the latent variable y_t^* follows an AR(2) model,

$$\begin{cases} y_t^* = \alpha + \beta y_{t-1}^* + \gamma y_{t-2}^* + \sigma \varepsilon_t^*, \\ y_t = \max(y_t^*, 0). \end{cases} \tag{26}$$

We obtain the same good results, as shown in Table 8.

	True	Starting Value	SLR Estimate	Standard Deviation	Mean Square Error
α	-0.3	0	-0.37491	0.070718	0.010613
β	0.2	-0.2	0.092988	0.088862	0.019348
σ	0.5	0.8	0.537858	0.043163	0.003296

Table 7: Tobit AR(1) model (25) - case 2 - MC replications 50 - acceptance rate 0.68 - zero percentage 0.7713.

	True	Starting Value	SLR Estimate	Standard Deviation	Mean Square Error
α	0	0.3	0.026658	0.066759	0.005167
β	-0.3	0.3	-0.26089	0.084696	0.008703
γ	0.1	0.24	0.107543	0.074104	0.005548
σ	0.8	1	0.777743	0.0482	0.002819

Table 8: Tobit AR(2) model (26) - MC replications 50 - acceptance rate 0.35 - zero percentage 0.5035.

4.3 Factor ARCH models

Multivariate ARCH models contain a large number of parameters and it is necessary to introduce some constraints in order to make this number smaller. A natural approach, compatible with the needs of financial theory and with some features of financial series which often have common evolution in the volatilities, leads to the introduction of unobserved factors (Diebold and Nerlove (1989), or Engle, Ng and Rothschild (1990)).

For simplicity and identifiability reasons, the original representation (16) is reduced to:

$$\begin{cases} y_t^* = (1+\beta y_{t-1}^{*2})^{1/2}\varepsilon_t^* & \varepsilon_t^* \sim \mathcal{N}(0,1), \\ y_t = a y_t^* + \varepsilon_t & \varepsilon_t \sim \mathcal{N}_p(0,\sigma^2 I_p), \end{cases} \tag{27}$$

and the dimension of y_t is $p=2$. As far as the estimation of this model is concerned Diebold and Nerlove (1989) propose to apply the extended Kalman filter, which leads to some approximations, while Gouriéroux *et al.* (1993) suggest the indirect inference approach.

The implementation of the Metropolis-Hastings simulation from

$$\tilde{f}(y^{*T}|y^T,\theta) \propto \exp\left\{-\sum_{t=1}^{T} ||y_t - a y_t^*||^2/2\sigma^2\right\} e^{-y_1^{*2}/2} \prod_{t=2}^{T} \frac{e^{-y_t^{*2}/2(1+\beta y_{t-1}^{*2})}}{(1+\beta y_{t-1}^{*2})^{1/2}}$$

relies on the first order importance function (see §3.3)

$$\tilde{f}(y_t^*|y_t, y_{t-1}^*,\theta) \propto \exp\left\{-\frac{||y_t - a y_t^*||^2}{2\sigma^2} - \frac{y_t^{*2}}{2(1+\beta y_{t-1}^{*2})}\right\},$$

which leads to the proposal distribution

$$\mathcal{N}\left(\frac{(1+\beta z_{t-1}^2)a' y_t}{(1+\beta z_{t-1}^2)||a||^2+\sigma^2}, \frac{(1+\beta z_{t-1}^2)\sigma^2}{(1+\beta z_{t-1}^2)||a||^2+\sigma^2}\right).$$

In order to illustrate the performance of this approach, we consider a simulated sample of size T=100 with true parameters $a=(-0.2,0.6)$, $\beta=0.8$ and $\sigma^2=0.2$. We consider both a global and a local approach, i.e. a global strategy of acceptance of the whole sequence z^T and a one move strategy (see §3.2). In the former case, the acceptance

	True	Starting Value	SLR Estimate
a_1	-0.2	-0.153	-0.14
a_2	0.6	0.43	0.42
b	0.8	0.86	0.99
σ^2	0.2	0.19	0.2

Table 9: Estimation result for the factor ARCH model (27) with a simulated sample of size $T = 100$ and a Bayes estimate as starting value.

probability is $\rho = \min(\omega, 1)$, with

$$\omega = \frac{\prod_{t=2}^{T} \frac{1}{\sqrt{(1+\bar\beta z_{t-1}^2)||\bar a||^2 + \bar\sigma^2}} \exp\left\{-\frac{1}{2} y_t'[(1+\bar\beta z_{t-1}^2)\bar a \bar a' + \bar\sigma^2 I_2]^{-1} y_t\right\}}{\prod_{t=2}^{T} \frac{1}{\sqrt{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2 + \bar\sigma^2}} \exp\left\{-\frac{1}{2} y_t'[(1+\bar\beta y_{t-1}^{*2})\bar a \bar a' + \bar\sigma^2 I_2]^{-1} y_t\right\}}$$

$$= \prod_{t=2}^{T} \frac{\sqrt{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2 + \bar\sigma^2}}{\sqrt{(1+\bar\beta z_{t-1}^2)||\bar a||^2 + \bar\sigma^2}} \exp\left\{-\frac{1}{2} \frac{(1+\bar\beta z_{t-1}^2) y_t' \bar a \bar a' y_t}{(1+\bar\beta z_{t-1}^2)||\bar a||^2 \bar\sigma^2 + \bar\sigma^4} + \frac{1}{2} \frac{(1+\bar\beta y_{t-1}^{*2}) y_t' \bar a \bar a' y_t}{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2 \bar\sigma^2 + \bar\sigma^4}\right\}$$

while in the latter it is $\rho_t = \min(\omega_t, 1)$, with

$$\omega_t = \frac{\tilde f(y_{t+1}^*(s\text{-}1)|z_t(s))\tilde f(z_t(s)|y_{t-1}^*(s\text{-}1))\tilde f(y_t|z_t(s))\tilde f(z_t(s)|y_{t-1}^*(s\text{-}1), y_t)}{\tilde f(y_{t+1}^*(s\text{-}1)|y_t^*(s\text{-}1))\tilde f(y_t^*(s\text{-}1)|y_{t-1}^*(s\text{-}1))\tilde f(y_t|y_t^*(s\text{-}1))\tilde f(y_t^*(s\text{-}1)|y_{t-1}^*(s\text{-}1), y_t)}$$

$$= \frac{\tilde f(y_{t+1}^*(s\text{-}1)|z_t(s))}{\tilde f(y_{t+1}^*(s\text{-}1)|y_t^*(s\text{-}1))} \sqrt{\frac{1+\bar\beta y_t^{*2}(s\text{-}1)}{1+\bar\beta z_t^2(s)}} \exp\left\{-\frac{1}{2}\frac{y_{t+1}^{*2}(s\text{-}1)}{1+\bar\beta z_t^2(s)} + \frac{1}{2}\frac{y_{t+1}^{*2}(s\text{-}1)}{1+\bar\beta y_t^{*2}(s\text{-}1)}\right\}$$

¿From the simulations, it appears that there is no difference between the two approaches in term of estimation results. We only consider the global approach in the following experience.

We found that the SLR method is quite sensitive in this case to the reference value, thus we choose to run the SLR algorithm with the noninformative Bayes estimate associated with the prior $\pi(\bar a, \bar\beta, \bar\sigma) = 1/\bar\sigma$. This estimate is obtained by a Metropolis-Hastings algorithm based on the same completion steps as the SLR algorithm, plus a conditional simulation of the parameters. In the particular simulation we run, the Bayes estimate is $\bar a = (-0.153, 0.43)$, $\bar\beta = 0.86$ and $(\bar\sigma)^2 = 0.19$. Figure 4.3 provides a slice representation of the estimated log-likelihood ratio, based on 1000 completed samples y^{*T}, recorded every 50 iterations, and the acceptance rate is 85-90%. The maximum is obtained at $\hat a = (-0.14, 0.42)$, $\hat\beta = 0.99$ and $\hat\sigma^2 = 0.2$ (see Table 9). The value of $\hat\beta$ is far from the true value β, but note that the estimated log-ratio at $(\hat a, \hat\beta, \hat\sigma)$ is 0.348, which indicates that the likelihood is quite flat in this region. The other phenomenon observed in this

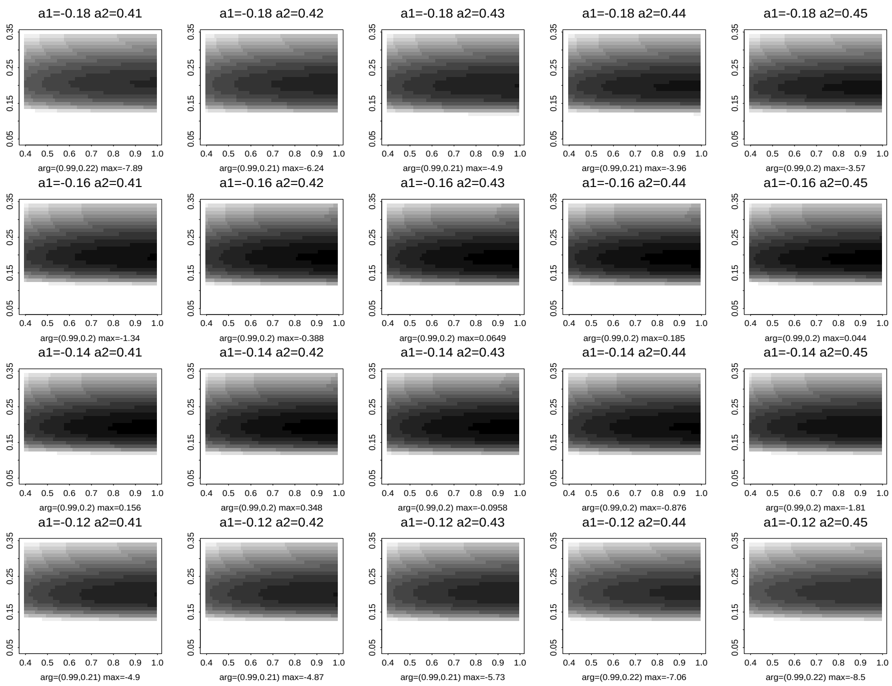

Figure 4: Estimated slices of the log-likelihood ratio for the factor ARCH model (27) for a simulated sample of size $T = 100$ and 50.000 iterations of the completion step, with a batch size of 50. The maximum (`max`) is achieved at $(\beta, \sigma^2) =$`arg` for each slice. See Table 9 for the parameter values.

setting is the shrinking estimation of a, which is systematically underestimated in the examples we tried.

4.4 Stochastic volatility models

Despite their intuitive appeal, stochastic volatility models have been used less frequently than ARCH models. This is mainly due to the difficulties associated with their estimation, since their likelihood function has no closed form. Existing estimation procedures can be subdivided into two groups: methods that attempt to build the likelihood function and methods which rely on an alternative, usually less efficient, approach. The quasi maximum likelihood method of Harvey, Ruiz and Shephard (1994), GMM methods of Andersen and Sorensen (1996) and the indirect inference approach of Monfardini (1998) are examples of the latter. Several propositions have also been made to evaluate the likelihood function. Kim, Shephard and Chib (1998) show how this function can be approximated when a mixture of normal distribution is used to describe the density of the disturbances. Jacquier *et al.* (1994) consider a Bayesian approach using a MCMC technique. Danielsson and Richard (1993) and Sandmann and Koopman (1997) suggest to approximate the likelihood function via importance sampling techniques. Finally Shephard (1993) propose a simulated EM approach.

Consider the stochastic volatility model (19). In the simulations we use a single move approach based on the proposals introduced in Section 3.6. We prefer the normal approach, since its acceptance rate is much higher (0.98) than both other candidate distributions (0.75 and 0.44 respectively). In particular, it is $\rho_t = \min(\omega_t, 1)$, with

$$\omega_t = \frac{\exp\left\{\frac{[z_t - \log(y_t^2)]^2}{4} - \frac{z_t}{2} - \frac{y_t^2 e^{-z_t}}{2}\right\}}{\exp\left\{\frac{[y_t^* - \log(y_t^2)]^2}{4} - \frac{y_t^*}{2} - \frac{y_t^2 e^{-y_t^*}}{2}\right\}}.$$

Moreover, the Bayes estimates used as reference values $\bar{\theta}$ are more satisfactory when using the normal approximation. We found indeed that the reference value of the SLR algorithm was crucial for the success or failure of the method in this example. When $\bar{\theta}$ was chosen at random and resulted in a value far away from the true value of θ, the maximum value of the approximation (11) was much closer to the reference value $\bar{\theta}$ than to the true value of θ. Replications of the method with solutions of the previous step as new reference values, as suggested in Geyer (1996), did not necessarily result in improved estimates, although the method usually converged to a fixed point. We therefore chose to run the SLR algorithm with a reasonable estimate and, as in the factor ARCH case, given that part of the algorithm was already completed, namely the simulation of the y_t^*'s, in order to obtain a Bayes estimate of the parameters, we only had to implement the second part of an MCMC algorithm, which was much easier given that the conditional distributions of $(\bar{a}, \bar{b})$ and $1/\bar{c}^2$ are normal and Gamma, respectively, with prior distribution $\pi(\bar{a}, \bar{b}, \bar{c}) = 1/\bar{c}$.

The Monte Carlo approximation of the likelihood ratio is based on the average of the following ratios:

$$\frac{f(y^T, y^{*T}(s); \theta)}{f(y^T, y^{*T}(s); \bar{\theta})} \propto \left(\frac{\bar{c}}{c}\right)^T \prod_{t=1}^{T} \frac{e^{-(y_t^*(s) - a - b y_{t-1}^*(s))^2 / 2c^2}}{e^{-(y_t^*(s) - \bar{a} - \bar{b} y_{t-1}^*(s))^2 / 2\bar{c}^2}},$$

since the observed part of the likelihood is parameter-free. This ratio factorizes through six statistics, since it is equal to

$$\left(\frac{\bar{c}}{c}\right)^T \exp\left[\frac{-1}{2}\left\{(c^{-2} - \bar{c}^{-2})\sum_{t=0}^{T}(y_t^*(s))^2 + (ac^{-2} - \bar{a}\bar{c}^{-2})(T+1) + (b^2 c^{-2} - \bar{b}^2 \bar{c}^{-2})\sum_{t=1}^{T}(y_{t-1}^*(s))^2 \right.\right.$$
$$+2(\bar{a}\bar{c}^{-2} - ac^{-2})\sum_{t=0}^{T} y_t^*(s) + 2(\bar{b}\bar{c}^{-2} - bc^{-2})\sum_{t=1}^{T} y_t^*(s) y_{t-1}^*(s)$$
$$\left.\left. +2(abc^{-2} - \bar{a}\bar{b}\bar{c}^{-2})\sum_{t=1}^{T} y_{t-1}^*(s)\right\}\right].$$

This property considerably reduces both computing time and storage problems.

Figure 4.4 illustrates the potential of the SLR method through a slice representation of the log-likelihood ratio in (a, b, c) and Table 10 presents the estimation result. It is important to point out that, as in the factor ARCH example, this estimation result have been obtained with a very small sample size ($T = 100$).

5 Conclusions

This paper proposes a MCMC approach in order to estimate the likelihood function and then to approximate the maximum likelihood estimator. Besides its generality,

	True	Starting value	SLR estimate
a	0.1	0.04819	0.07
b	0.8	0.9063	0.88
c	0.1	0.05701	0.12

Table 10: Estimation result for the stochastic volatility model (19) with a simulated sample of size $T = 100$ and a Bayes estimate as starting value.

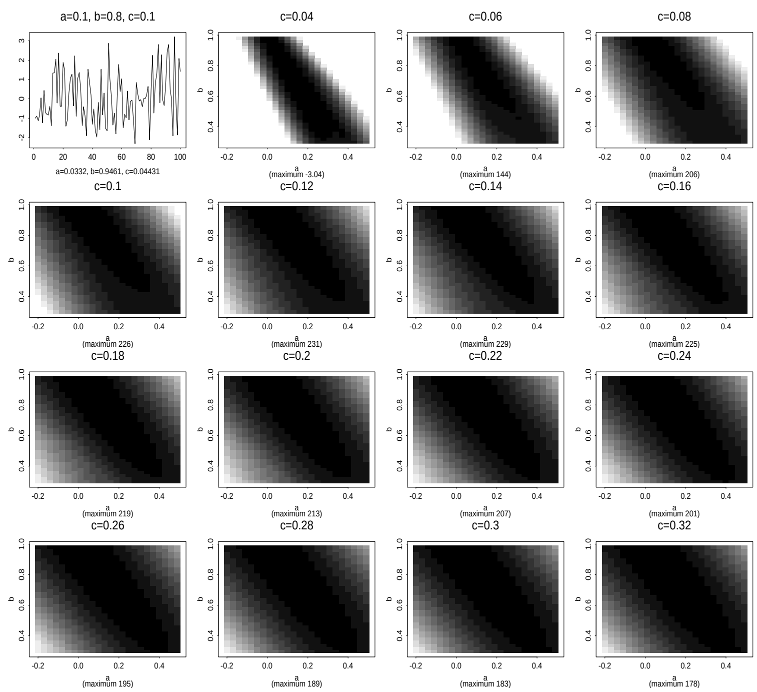

Figure 5: Evaluated slices in c of the log-likelihood ratio associated with (19) for a simulated sample of size $T = 100$ and 5000 iterations of the completion step. See Table 10 for the parameter values.

since it applies to most latent variables models, this method presents some important advantages. First, provided the number of simulation is large enough, the Simulated Likelihood Ratio method can be considered as asymptotically efficient, which is not the case for the methods of simulated moments, simulated-pseudo likelihood methods or indirect inference methods. Second, and contrary to simulated EM methods, the SLR method only requires one optimization (and not a sequence of optimizations) and it allows for the computation of the likelihood surface and, for instance, likelihood ratio test statistics. Third, the simulation is made for only one value of parameters and, therefore, the objective function will always be smooth with respect to them, even if the simulation involves rejection methods or if the endogenous variables are non differentiable or discontinuous function of the parameters. Moreover, in some situations, when the simulation is time consuming, it may be interesting to store the simulated paths, instead of redrawing them, when computing the likelihood ratio for different values of parameters. Finally, the results obtained in several numerical applications are quite satisfactory, although the influence of the reference value may be important in some examples.

6 Appendix: Theoretical developments

6.1 A formal study of conditionals

Although this may appear over-technical, it is necessary to set the basis for the importance sampling techniques in the most rigorous manner. We thus assume that the vector (y^{*T}, y^T) is absolutely continuous, i.e. has a density, w.r.t. the product measure $\otimes_{t=1}^T \nu_t$, where the ν_t's are σ-finite positive measures on $\mathrm{R}^{p^*} \times \mathrm{R}^p$, endowed with its Borel σ-algebra. We denote by $f_0(y^{*T}, y^T)$ this unknown p.d.f., which is assumed to belong to a parametric family,

$$\{f(y^{*T}, y^T; \theta),\ \theta \in \Theta \subset \mathrm{R}^k\}.$$

These assumptions imply that $f(y^{*T}, y^T; \theta)$ can be factorized into a product of conditional p.d.f.'s, w.r.t. the ν_t's,

$$\begin{aligned} f(y^{*T}, y^T; \theta) &= \prod_{t=1}^T f_t(y_t^*, y_t | y^{*t-1}, y^{t-1}; \theta) \\ &= \prod_{t=1}^T f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta), \end{aligned} \tag{28}$$

where, by convention, $f_t(y_1^*, y_1 | y^{*0}, y^0; \theta) = f(y_1^*, y_1; \theta)$ is the unconditional p.d.f. of (y_1^*, y_1). We also assume, without loss of generality, that the measures ν_t can be disintegrated in the following way: for every Borel sets $A \subset \mathrm{R}^{p^*}$ and $B \subset \mathrm{R}^p$,

$$\nu_t(A \times B) = \int_A \mu_t^{y^*}(B) d\mu_{*t}(y^*) = \int_B \mu_{*t}^{y}(A) d\mu_t(y)\,,$$

where $\mu_t^{\cdot} = \{\mu_t^{y^*}, y^* \in \mathrm{R}^{p^*}\}$ (respectively $\mu_{*t}^{\cdot} = \{\mu_{*t}^{y}, y \in \mathrm{R}^p\}$) is a set of σ-finite measures on R^p (respectively R^{p^*}) indexed by $y^* \in \mathrm{R}^{p^*}$ (respectively $y \in \mathrm{R}^p$) and μ_{*t} (respectively μ_t) is a σ-finite measure on R^{p^*} (respectively R^p). These disintegrations will be denoted more concisely as

$$\nu_t = \mu_{*t} \otimes \mu_t^{\cdot} = \mu_{*t}^{\cdot} \otimes \mu_t\,.$$

In this framework, the conditional distributions of y_t^* and y_t, given $\mathcal{I}_{t-1}$, have a p.d.f. w.r.t. μ_{*t} and μ_t given by

$$\begin{aligned} f_t(y_t^*|\mathcal{I}_{t-1};\theta) &= \int f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta) d\mu_t^{y_t^*}(y_t)\,, \\ f_t(y_t|\mathcal{I}_{t-1};\theta) &= \int f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta) d\mu_{*t}^{y_t}(y_t^*)\,, \end{aligned}$$

respectively. Similarly, the conditional distributions of y_t^* given $(\mathcal{I}_{t-1}, y_t)$ and of y_t given $(\mathcal{I}_{t-1}, y_t^*)$ have p.d.f. w.r.t. $\mu_{*t}^{y_t}$ and $\mu_t^{y_t^*}$ given by

$$\begin{aligned} f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) &= \frac{f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta)}{f_t(y_t|\mathcal{I}_{t-1};\theta)}\,, \\ f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta) &= \frac{f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta)}{f_t(y_t^*|\mathcal{I}_{t-1};\theta)}\,, \end{aligned} \tag{29}$$

respectively. Using (29), the global p.d.f. $f_t(y^{*T}, y^T;\theta)$ given in (28) can be further factorized as

$$\begin{aligned} f(y^{*T}, y^T;\theta) &= \prod_{t=1}^{T} f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) f_t(y_t|\mathcal{I}_{t-1};\theta) \\ &= \prod_{t=1}^{T} f_t(y_t^*|\mathcal{I}_{t-1};\theta) f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta)\,. \end{aligned} \tag{30}$$

The following result shows that Bayes' Theorem also applies for the derivation of conditional distributions in this setup:

Theorem 6.1 *The random variables y^{*T} and y^T have p.d.f.'s w.r.t. $\otimes_{t=1}^T \mu_{*t}$ and $\otimes_{t=1}^T \mu_t$ given by:*

$$\begin{aligned} f(y^{*T};\theta) &= \int f(y^{*T}, y^T;\theta) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t), \\ f(y^T;\theta) &= \int f(y^{*T}, y^T;\theta) \otimes_{t=1}^T d\mu_{*t}^{y_t}(y_t^*), \end{aligned} \tag{31}$$

and the conditional distributions of y^{*T} given y^T and of y^T given y^{*T} have p.d.f. w.r.t. $\otimes_{t=1}^T d\mu_{*t}^{y_t}$ and $\otimes_{t=1}^T d\mu_t^{y_t^*}$ given by

$$\begin{aligned} f(y^{*T}|y^T;\theta) &= \frac{f(y^{*T}, y^T;\theta)}{f(y^T;\theta)}, \\ f(y^T|y^{*T};\theta) &= \frac{f(y^{*T}, y^T;\theta)}{f(y^{*T};\theta)}, \end{aligned}$$

respectively .

This result follows from the following technical lemma:

Lemma 6.2 *Let $h(y^{*T}, y^T)$ be a real, measurable, positive or $\otimes_{t=1}^T \nu_t$ integrable function. The integral*

$$I(h) = \int h(y^{*T}, y^T) \otimes_{t=1}^T d\nu_t(y_t^*, y_t)$$

is equal to

$$\int \left[\int h(y^{*T}, y^T) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \right] \otimes_{t=1}^T d\mu_{*t}(y_t^*).$$

Proof of Lemma 6.2. $I(h)$ can be computed by successive integrations w.r.t. the measures $d\mu_1^{y_1^*}(y_1)$, $d\mu_{*1}(y_1^*)$, $d\mu_2^{y_2^*}(y_2)$, $d\mu_{*2}(y_2^*)$, $d\mu_3^{y_3^*}(y_3)$, $d\mu_{*3}(y_3^*)$, ... , $d\mu_T^{y_T^*}(y_T)$, $d\mu_{*T}(y_T^*)$. Once the integration w.r.t. $d\mu_1^{y_1^*}(y_1)$ is done, we have to integrate the function thus obtained w.r.t. the product measure $d\mu_{*1}(y_1^*) \otimes d\mu_2^{y_2^*}(y_2)$. Fubini's Theorem states that the order of integration can be inverted and, integrating first w.r.t. $d\mu_2^{y_2^*}(y_2)$, we get a function which must be integrated w.r.t. the product measure $d\mu_{*1}(y_1^*) \otimes d\mu_{*2}(y_2^*) \otimes d\mu_3^{y_3^*}(y_3)$. Using Fubini's Theorem once more, we can first integrate w.r.t. $d\mu_3^{y_3^*}(y_3)$. Iterating this procedure, we see that $I(h)$ can be written

$$I(h) = \int \left[\int h(y^{*T}, y^T) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \right] \otimes_{t=1}^T d\mu_{*t}(y_t^*) \,,$$

which establishes Lemma 1.

Consider a Borel set A of R^{Tp^*}. We have

$$\begin{aligned} P_\theta^{y^{*T}}(A) &= P_\theta^{(y^{*T}, y^T)}(A \times \mathrm{R}^{Tp}) \\ &= \int 1\!\!1_{A \times \mathrm{R}^{Tp}}(y^{*T}, y^T) f(y^{*T}, y^T; \theta) \otimes_{t=1}^T d\nu_t(y_t^*, y_t), \end{aligned}$$

where P^Z is the distribution of Z. Using Lemma 1, we get

$$P_\theta^{y^{*T}}(A) = \int_A \left[\int f(y^{*T}, y^T; \theta) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \right] \otimes_{t=1}^T d\mu_{*t}(y_t^*) \,.$$

Thus, $P_\theta^{y^{*T}}$ is absolutely continuous w.r.t. $\otimes_{t=1}^T \mu_{*t}(y_t^*)$, with the p.d.f.

$$f(y^{*T}; \theta) = \int f(y^{*T}, y^T; \theta) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \,.$$

Similarly,

$$f(y^T; \theta) = \int f(y^{*T}, y^T; \theta) \otimes_{t=1}^T d\mu_{*t}^{y_t}(y_t^*),$$

and the formula (31) is established.

Now let A and B be Borel sets of R^{p^*T} and R^{pT}, respectively. Lemma 1 implies

$$\begin{aligned} P_\theta^{y^{*T}, y^T}(A \times B) &= \int_A \left[\int_B f(y^{*T}, y^T; \theta) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \right] \otimes_{t=1}^T d\mu_{*t}(y_t^*) \\ &= \int_A \left[\int_B \frac{f(y^{*T}, y^T; \theta)}{f(y^{*T}; \theta)} \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t) \right] dP_\theta^{y^{*T}}(y^{*T}) \,. \end{aligned}$$

This shows that the conditional distribution of y^T given y^{*T} is absolutely continuous w.r.t. the measure $\otimes_{t=1}^T \mu_t^{y_t^*}(y_t)$ and that its p.d.f. is $f(y^{*T}, y^T; \theta)/f(y^{*T}; \theta)$. The symmetric result is obvious and the formulas (3) are thus established.

6.2 Proof of Theorem 3.1

Note first that Lemma 1 implies that, for any Borel set $A_k \subset \mathrm{R}^{p^*(T-k)}$, $B \subset \mathrm{R}^{pT}$,

$$
\begin{aligned}
P_\theta^{y^{*T-k},y^T}(A_k \times B) &= \int_B \left[\int_{A_k\times \mathrm{R}^{p^*k}} f(y^{*T}, y^T;\theta) \otimes_{t=1}^T d\mu_{*t}^{y_t}(y_t^*)\right] \otimes_{t=1}^T d\mu_t(y_t) \\
&= \int_B \left[\int_{A_k} \left[\int f(y^{*T}, y^T;\theta) \otimes_{t=k+1}^T d\mu_{*t}^{y_t}(y_t^*)\right] \otimes_{t=1}^{T-k} d\mu_{*t}^{y_t}(y_t^*)\right] \otimes_{t=1}^T d\mu_t(y_t)\,.
\end{aligned}
$$

This shows that $P_\theta^{y^{*T-k},y^T}(A_k \times B)$ is absolutely continuous w.r.t. $\otimes_{t=1}^{T-k}\nu_t \otimes_{t=T-k+1}^T \mu_t$ and has a p.d.f. given by

$$
f(y^{*T-k}, y^T;\theta) = \int f(y^{*T}, y^T;\theta) \otimes_{t=k+1}^T d\mu_{*t}^{y_t}(y_t^*)\,.
$$

We also have

$$
f(y^{*T-k}|y^T;\theta) = \int f(y^{*T}|y^T;\theta) \otimes_{t=k+1}^T d\mu_{*t}^{y_t}(y_t^*) = \frac{f(y^{*T-k}, y^T;\theta)}{f(y^T;\theta)}\,.
$$

Moreover, the expectation of Theorem 3.1 can be written

$$
\begin{aligned}
E_{\bar\theta}\left[\frac{f(y^{*T-k}, y^T;\theta)}{f(y^{*T-k}, y^T;\bar\theta)}\Big| y^T\right] &= \int \frac{f(y^{*T-k}, y^T;\theta)}{f(y^{*T-k}, y^T;\bar\theta)} \frac{f(y^{*T-k}, y^T;\bar\theta)}{f(y^T;\bar\theta)} \otimes_{t=1}^{T-k} d\mu_{*t}^{y_t}(y_t^*) \\
&= \frac{1}{f(y^T;\bar\theta)} \int f(y^{*T-k}, y^T;\theta) \otimes_{t=1}^{T-k} d\mu_{*t}^{y_t}(y_t^*) \\
&= \frac{f(y^T;\theta)}{f(y^T;\bar\theta)}\,.
\end{aligned}
$$

6.3 Conditional computation

The framework of this paper allows for the computation of various conditional p.d.f.'s. For instance, the conditional distribution of y_t^* given (y^{*t-1}, y^{t+1}) has a p.d.f. w.r.t. $\mu_{*t}^{y_t}$,

$$
f_t(y_t^*|y^{*t-1}, y^{t+1};\theta) = \frac{f_{t+1}(y_{t+1}|y^{*t}, y^t;\theta) f_t(y_t^*|y^{*t-1}, y^t;\theta)}{\int f_{t+1}(y_{t+1}|y^{*t}, y^t;\theta) f_t(y_t^*|y^{*t-1}, y^t;\theta) d\mu_{t+1}(y_{t+1})}\,.
$$

First, using arguments similar to Appendix 6.1, we can see that the distribution of (y^{*t-1}, y^{t+1}) has a p.d.f. w.r.t. $\otimes_{s=1}^{t-1}\nu_s \otimes \mu_t \otimes \mu_{t+1}$,

$$
f(y^{*t-1}, y^{t+1};\theta) = \int f(y^{*T}, y^T;\theta) \otimes_{s=t+2}^T d\nu_s \otimes d\mu_{*t}^{y_t} \otimes d\mu_{*t+1}^{y_{t+1}}\,.
$$

Therefore, for any Borel sets $A \subset \mathrm{R}^{p^*}$, $B \subset \mathrm{R}^{p^*(t-1)} \times \mathrm{R}^{p(t+1)}$,

$$\begin{aligned}
P_\theta^{y_t^*, y^{*t-1}, y^{t+1}}(A \times B) &= \int \mathbb{1}_B(y^{*t-1}, y^{t+1}) \left[\int \mathbb{1}_A(y_t^*) \left[\int f(y^{*T}, y^T; \theta) \right.\right. \\
&\qquad \left.\left. \otimes_{s=t+2}^T d\nu_s \otimes d\mu_{*t+1}^{y_{t+1}}\right] \otimes d\mu_{*t}^{y_t}\right] \otimes_{s=1}^{t-1} d\nu_s \otimes d\mu_t \otimes d\mu_{t+1} \\
&= \int_B \left[\int_A \frac{f(y^{*t}, y^{t+1}; \theta)}{f(y^{*t-1}, y^{t+1}; \theta)} d\mu_{*t}^{y_t}\right] f(y^{*t-1}, y^{t+1}; \theta) \otimes_{s=1}^{t-1} d\nu_s \otimes d\mu_t \otimes d\mu_{t+1} \\
&= \int_B \left[\int_A \frac{f(y^{*t}, y^{t+1}; \theta)}{f(y^{*t-1}, y^{t+1}; \theta)} d\mu_{*t}^{y_t}\right] dP_\theta^{(y^{*t-1}, y^{t+1})},
\end{aligned}$$

and, therefore, the conditional distribution of y_t^* given (y^{*t-1}, y^{t+1}) has a p.d.f. w.r.t. $\mu_{*t}^{y_t}$ which is $f(y^{*t}, y^{t+1}; \theta)/f(y^{*t-1}, y^{t+1}; \theta) = f_t(y_t^* | y^{*t-1}, y^{t+1}; \theta)$, say. ¿From the form of these joint p.d.f.'s, we get

$$\begin{aligned}
f_t(y_t^* | y^{*t-1}, y^{t+1}; \theta) &= \frac{\int f(y^{*T}, y^T; \theta) \otimes_{s=t+2}^T d\nu_s \otimes d\mu_{*t+1}^{y_{t+1}}}{\int f(y^{*T}, y^T; \theta) \otimes_{s=t+2}^T d\nu_s \otimes d\mu_{*t}^{y_t} \otimes d\mu_{*t+1}^{y_{t+1}}} \\
&= \frac{\int f(y^{*t+1}, y^{t+1}; \theta) d\mu_{*t+1}^{y_{t+1}}}{\int f(y^{*t+1}, y^{t+1}; \theta) d\mu_{*t}^{y_t} \otimes d\mu_{*t+1}^{y_{t+1}}} \\
&= \frac{f(y^{*t}, y^t; \theta) \int f_{t+1}(y_{t+1}^*, y_{t+1} | y^{*t}, y^t; \theta) d\mu_{*t+1}^{y_{t+1}}}{f(y^{*t-1}, y^{t-1}; \theta) \int f_t(y_{t+1}^*, y_{t+1}, y_t^*, y_t | y^{*t-1}, y^{t-1}; \theta) d\mu_{*t}^{y_t} \otimes d\mu_{*t+1}^{y_{t+1}}} \\
&= \frac{f_t(y_t^*, y_t | y^{*t-1}, y^{t-1}; \theta) f_{t+1}(y_{t+1} | y^{*t}, y^t; \theta)}{\int f_{t+1}(y_{t+1} | y^{*t}, y^t; \theta) f_t(y_t^*, y_t | y^{*t-1}, y^{t-1}; \theta) d\mu_{*t}^{y_t}} \\
&= \frac{f_t(y_t^* | y^{*t-1}, y^t; \theta) f_{t+1}(y_{t+1} | y^{*t}, y^t; \theta)}{\int f_t(y_t^* | y^{*t-1}, y^t; \theta) f_{t+1}(y_{t+1} | y^{*t}, y^t; \theta) d\mu_{*t}^{y_t}},
\end{aligned}$$

simplifying by $f_t(y_t | y^{*t-1}, y^{t-1}; \theta)$.

References

Andersen, T.G. and Sorensen, B.E. (1996), GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study, *J. Business & Economic Statistics*, 14, 328–352.

Billio, M. and Monfort, A. (1998), Switching State Space Models, *J. Statist. Plann. Inf.*, 68, 65–103.

Billio, M., Monfort, A. and Robert C.P. (1999), Bayesian Estimation of Switching ARMA Models, *J. Econometrics* **93**, 229–255.

Celeux, G. and Diebolt, J. (1985), The SEM Algorithm: a Probabilistic Teacher derived from the EM Algorithm for the Mixture problem. *Comp. Statist. Quater.*, 2, 73–82.

Danielsson, J. and Richard, J.F. (1993), Accelerated Gaussian Importance Sampler with Application do Dynamic Latent Variable Models, *J. Applied Econom.*, 8, 153–173.

Diebold, D. and Nerlove, M. (1989), The Dynamic of Exchange Rate Volatility: a Multivariate Latent Factor ARCH Model, *J. Applied Econom.*, 4, 1–22.

Diebolt, J. and Ip, E.H.S. (1996), Stochastic EM: Method and Applications, in *Markov Chain Monte Carlo in Practice*, eds Gilks W.R., Richardson S. and Spiegelhalter D.J., Chapman and Hall, London.

Dueker, M.J. (1997), Markov Switching in GARCH Processes and Mean-Reverting Stock Market Volatility, *J. Business & Economic Stat.*, 15, 26–34.

Duffie, J. and Singleton, K. (1993), Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929–952.

Engle, R., Ng, V. and Rothschild, M. (1990), Asset Pricing with a Factor ARCH Variance Structure: Empirical Estimates for Treasury Bills, *J. Econometrics*, 45, 213–237.

Geyer, C.J. (1994), On the Convergence of Monte Carlo Maximum Likelihood Calculations, *J. Royal Statist. Soc.* (Ser. B) 56, 261–274.

Geyer, C.J. (1996), Estimation and Optimization of Functions, in *Markov Chain Monte Carlo in Practice*, eds Gilks W.R., Richardson S. and Spiegelhalter D.J., Chapman and Hall, London.

Geyer, C.J. and Thompson, E.A. (1992), Constrained Monte Carlo Maximum Likelihood for Dependent Data (with discussion). *J. Royal Statist. Soc.* (Ser. B), 54, 657–699.

Geweke, J. (1991), Efficient Simulation from the Multivariate Normal and Student's t-Distributions subject to Linear Constraints, *Computer Sciences and Statistics: Proc. 23d Symp. Interface*, 571–577.

Ghysels, E., Harvey A.C. and Renault, E. (1996), Stochastic Volatility, in *Handbook of Statistics*, vol. 14 *Statistical Methods in Finance*, 5, 119-191, eds C.R. Rao and G.S. Maddala, Elsevier, Amsterdam.

Gouriéroux, C. and Monfort, A. (1996), *Simulation Based Methods in Econometrics*, Oxford University Press, Oxford.

Gouriéroux, C., Monfort, A. and Renault, E. (1993), Indirect Inference, *J. Applied Econom.*, 8, 85–118.

Hamilton, J.D. (1988), Rational-Expectations Econometric Analysis of Changes in Regime: an Investigation of the Term Structure of Interest Rates, *J. Economic Dynamics and Control*, 12, 385–423.

Hamilton, J.D., A new Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica*, 57, 357–384, 1989.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994), Multivariate Stochastic Variance Models, *Review of Economic Studies* 61, 247–264.

Jacquier, E., Polson, N.G. and Rossi, P.E. (1994), Bayesian Analysis of Stochastic Volatility Models, *J. Business & Economic Stat.*, 12, 371–417.

Kim, C.J. (1994), Dynamic Linear Models With Markov Switching, *J. Econometrics* , 64, 1–22.

Kim, S., Shephard, N. and Chib, S. (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, *Review of Economic Studies*, 65/3 , 361–394.

Lam, P.S. (1990), The Hamilton Model with a General Autoregressive Component, *J. Monetary Economics*, 26, 409-432.

Laroque, G. and Salanié, B. (1993), Simulation Based Estimation of Models with Lagged Latent Variables, *J. Applied Econom.*, 8, 119–133.

Lee, L.F. (1997a), Simulation Estimation of Dynamic Switching Regression and Dynamic Disequilibrium Models - some Monte Carlo Results, *J. Econometrics*, 78, 179–204.

Lee, L.F. (1997b), A Smooth Likelihood Simulator for Dynamic Disequilibrium Models, *J. Econometrics*, 78, 257–294.

Manrique, A. and Shephard, N. (1998), Simulation based Likelihood Inference for Limited Dependent Processes, *Econom. Journal*, 1, 174–202.

McFadden, D. (1989), A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, *Econometrica*, 57, 995–1026.

Monfardini, C. (1998), Estimating Stochastic Volatility Models through Indirect Inference, *Econom. Journal*, 1, 113–128.

Philippe, A. (1997), Simulation of Right- and Left-truncated Gamma Distributions by Mixtures, *Statistics & Computing*, 7, 173–182.

Robert, C.P. and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Robert, C.P. and Titterington, M. (1998), Resampling Schemes for Hidden Markov Models and their Application for Maximum Likelihood Estimation, *Statistics and Computing* 8, 145–158.

Sandmann, G. and Koopman, S.J. (1997), Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood, *mimeo*, London School of Economics.

Shephard, N. (1993), Fitting Non-linear Time Series Models, with Applications to Stochastic Variance Models, *J, Applied Econom.*, 8, 135–152.

Shephard, N. (1994), Partial Non-Gaussian State Space, *Biometrika*, 81, 115–131.

Shephard, N. (1996), Statistical Aspects of ARCH and Stochastic Volatility Models, in *Time Series Models*, eds D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, Chapman and Hall, London.

Shephard, N. and Pitt, M.K. (1997), Likelihood Analysis of Non-Gaussian Measurement Time Series, *Biometrika*, 84, 653–668.

Wei, G.C.G. and Tanner, M.A. (1990), A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithm, *J. American Statist. Assoc.*, 85, 699–704.