THE SIMULATED LIKELIHOOD RATIO (SLR) METHOD

M. BILLIO[1]
A. MONFORT[2]
C.P. ROBERT[3]

March 1998

Résumé. Une méthode de simulation fondée sur des techniques d'echantillonage d'importance, de Gibbs et Metropolis-Hastings permet d'approcher le rapport de la vraisemblance en deux valeurs du paramètre, et donc l'estimateur du maximum de vraisemblance, dans le cadre général des modèles dynamiques avec variables latentes. Cette classe contient comme cas particuliers les modèles à facteurs, les modèles à changements de régimes, les modèles à variable dépendante limitée dynamiques, les modèles à volatilité stochastique et les modèles en temps continu discrétisés.

Abstract. A simulation method based on importance sampling, Gibbs and Metropolis-Hastings techniques allows to approximate the ratio between the likelihood function computed for two different parameter values. Thus it is possible to approximate the maximum likelihood estimator in the general framework of dynamic latent variable models. Some examples of this class of models are factor models, switching regime models, dynamic limited dependent variable models, stochastic volatility models and discretised continuous time models.

[1]University Ca' Foscari of Venice
[2]CREST-INSEE
[3]CREST-INSEE

1 INTRODUCTION

The class of *parametric dynamic latent variable models* (also called factor models or hidden variable models or hierarchical models) is becoming increasingly popular, because of the flexibility they offer in the modelling of complex phenomena. They 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 $f(y_1, \cdots, y_T; \theta)$, which requires a high dimensional integral, whose size is equal to the number of observations multiplied by the size of the unobserved variables. Simple examples of these models include *hidden Markov mixture models*, where the unobserved y_t^*'s carry the time dependence, the y_t's being independent conditionally on the y_t^*'s, or *censored AR models* where the y_t^*'s are distributed from a standard AR(p) model and $y_t = \min\{y_t^*, x_t\}$. More involved setups considered in this paper include factor models, switching regime dynamic models, stochastic volatility models and discretized continuous time models. These models are commonly used, not only in Econometrics (see Gouriéroux and Monfort [20]) but also in Finance (Jacquier, Polson and Rossi [27]) and survey theory (Little and Rubin [34], Heitjan and Rubin [26]).

Simulation based inference methods propose several ways of tackling the inference problem for this class of models. A first approach relies on simulation based 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 [38], Duffie and Singleton [10]), the simulated pseudo-maximum likelihood method (Laroque and Salanié [31]), the indirect inference method (Gouriéroux, Monfort and Renault [21]). A second approach considers the problem of the computation (or of the approximation) of the maximum likelihood estimator through importance sampling methods (Danielsson and Richard [6], Lee [32], Billio and Monfort [3]) or Simulated EM (Shephard [44], [45], Diebolt and Ip [8]). A third approach borrows from Bayesian MCMC algorithms (Albert and Chib [2], Jacquier, Polson and Rossi [27], Robert and Titterington [42], Billio, Monfort and Robert [4]).

This paper proposes a fourth approach which utilizes MCMC methods in order to compute (or approximate) the maximum likelihood estimator (see also Geyer [13], [14], Geyer and Thomson [15]).
The first step of this approach is the building of a precise probability setup which carefully defines the relevant marginal or conditional probability density functions (p.d.f.) and the corresponding dominating measures; this fairly general setup contains, as particular cases, all the usual dynamic latent variables models, including the degenerate cases where the observable variables are deterministic functions of the latent variables. The second and main step of our approach is to establish that the 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 value of the parameter, can be expressed as the conditional expectation

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

where k is any integer. A key feature in this representation is that the likelihood ratio thus appears as the expectation of a completed likelihood ratio with respect to a *single distribution*, namely the conditional distribution of the y_t^*'s given the y_t's, corresponding to a fixed value $\bar{\theta}$ of the parameter. This property has important bearings on the implementation of the method.
The third step exploits this property and proposes to evaluate this ratio by a Monte Carlo method. The simulations in the appropriate distribution, namely the distribution of $y_1^*, \cdots, y_{T-k}^*$ conditional on $y_1, \cdots, y_T$, when $\theta = \bar{\theta}$, are performed by using a Metropolis-Hastings procedure and, given the fact that this distribution is fixed, the objective function is smooth in the parameter θ even when observable variables are non differentiable or non continuous functions of the parameter, as it is the case respectively in dynamic disequilibrium models, or in switching state space models (see Lee [32],[33]). Moreover, since the simulations are done with a fixed distribution, the simulation step is only called once (in an iterative procedure maximizing the likelihood function) and can be used in a vectorialized procedure computing simultaneously the likelihood ratio at many values of the parameters (likelihood surfaces or sections are thus easily computed).
The fourth step of the procedure is to choose an instrumental distribution for the Metropolis-Hastings procedure. Several proposals are made, including Sequential Optimal Sampling distributions proposed in Billio and Monfort [3]. Finally the likelihood ratio thus obtained is maximized, providing approximation of asymptotically efficient estimators and, if desired, of likelihood ratio test statistics.

The paper is organised as follows. Section 2 describes the general setup providing a rigorous approach to the corresponding probabilistic models and deriving the necessary conditional distribution. Section 3 introduces the general Simulated Likelihood Ratio (SLR) procedure and shows how it applies to various classical models. Section 4 provides several numerical applications. Concluding remarks are given in Section 5 and the proofs are gathered in three appendices.

2 THE SETUP

2.1 Basic assumptions

Although this first section may appear over-technical, it is necessary to set the basis for the following importance sampling techniques in the most rigorous manner.
We consider two sets of vectors: the y_t^*, $t = 1, ..., T$, are p^*-dimensional vectors of possibly unobservable endogenous variables and the y_t, $t = 1, ..., T$, are p-dimensional vectors of observable endogenous variables. For simplicity's sake exogenous variables are not mentioned explicitly, although they can be introduced at no cost.
We adopt the following notations: $y^{*t} = (y_1^*, ..., y_t^*)$, $y^t = (y_1, ..., y_t)$, and $\mathcal{I}_t = \{y^{*t}, y^t\}$. We assume that (y^{*T}, y^T) is absolutely continuous, i.e. has a density, with respect to the product measure $\otimes_{t=1}^T \nu_t$, where the ν_t's are σ-finite positive measures on $\mathbb{R}^{p^*} \times \mathbb{R}^p$, endowed with their Borel σ-algebra. We denote by $f_0(y^{*T}, y^T)$ this unknown probability density function (p.d.f.) and we assume that it belongs to the parametric family:

$$\{f(y^{*T}, y^T; \theta),\ \theta \in \Theta \subset \mathbb{R}^k\}. \tag{1}$$

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

$$\begin{aligned} f(y^{*T}, y^T; \theta) &= \prod_{t=1}^{T} \ f_t(y_t^*, y_t | y^{*t-1}, y^{t-1}; \theta) \\ &= \prod_{t=1}^{T} \ f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta), \end{aligned} \tag{2}$$

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's can be disintegrated in the following way: for every Borel sets $A \subset I\!\!R^{p^*}$ and $B \subset I\!\!R^p$,

$$\nu_t(A \times B) = \int_A \mu_t^{y^*}(B) d\mu_{*t}(y^*) = \int_B \mu_{*t}^{y}(A) d\mu_t(y) \tag{3}$$

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

$$\nu_t = \mu_{*t} \otimes \mu_t^{\cdot} = \mu_{*t}^{\cdot} \otimes \mu_t \tag{4}$$

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

$$\begin{aligned} f_t(y_t^* | \mathcal{I}_{t-1}; \theta) &= \int f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta) d\mu_t^{y_t^*}(y_t), \\ f_t(y_t | \mathcal{I}_{t-1}; \theta) &= \int f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta) d\mu_{*t}^{y_t}(y_t^*), \end{aligned} \tag{5}$$

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. with respect to $\mu_{*t}^{y_t}$ and $\mu_t^{y_t^*}$ given by:

$$\begin{aligned} f_t(y_t^* | \mathcal{I}_{t-1}, y_t; \theta) &= \frac{f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta)}{f_t(y_t | \mathcal{I}_{t-1}; \theta)}, \\ f_t(y_t | \mathcal{I}_{t-1}, y_t^*; \theta) &= \frac{f_t(y_t^*, y_t | \mathcal{I}_{t-1}; \theta)}{f_t(y_t^* | \mathcal{I}_{t-1}; \theta)}, \end{aligned} \tag{6}$$

respectively. Using (6), the global p.d.f. $f_t(y^{*T}, y^T; \theta)$ given in (2) can be further factorized as:

$$\begin{aligned} f(y^{*T}, y^T; \theta) &= \prod_{t=1}^{T} f_t(y_t^* | \mathcal{I}_t, y_t; \theta) f_t(y_t | \mathcal{I}_t; \theta) \\ &= \prod_{t=1}^{T} f_t(y_t^* | \mathcal{I}_t; \theta) f_t(y_t | \mathcal{I}_t, y_t^*; \theta) \end{aligned} \tag{7}$$

The following proposition will be useful in the derivation of conditional distributions:

Proposition 1
The random variables y^{*T} and y^T have p.d.f.'s with respect to $\otimes_{t=1}^T \mu_{*t}$ and $\otimes_{t=1}^T \mu_t$ given by:

$$\begin{aligned} f(y^{*T};\theta) &= \int f(y^{*T}, y^T;\theta) \otimes_{t=1}^T d\mu_t^{y_t^*}(y_t), \\ f(y^T;\theta) &= \int f(y^{*T}, y^T;\theta) \otimes_{t=1}^T d\mu_{*t}^{y_t}(y_t^*), \end{aligned} \tag{8}$$

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

$$\begin{aligned} f(y^{*T}|y^T;\theta) &= \frac{f(y^{*T}, y^T;\theta)}{f(y^T;\theta)}, \\ f(y^T|y^{*T};\theta) &= \frac{f(y^{*T}, y^T;\theta)}{f(y^{*T};\theta)}, \end{aligned} \tag{9}$$

respectively .

Proof: see Appendix 1.

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

2.2 The "product" latent models

In this case, all the measures in the set $\{\mu_t^{y^*}, y^* \in I\!R^{p^*}\}$ are identical, and denoted by μ_t, or, equivalently, all the measures in $\{\mu_{*t}^y, y \in I\!R^p\}$ are identical, and denoted by μ_{*t}. In such a case ν_t is simply a product measure:

$$\nu_t = \mu_{*t} \otimes \mu_t \tag{10}$$

This means, for instance, that all the conditional probabilities of y_t given $(\mathcal{I}_{t-1}, y_t^*)$ have a p.d.f. with respect to the same measure μ_t (and similarly for y_t^*).
Such a case typically occurs when (y_t^*, y_t) satisfies a model of the following type:

$$\left\{ \begin{array}{ll} y_t^* & = r_t^*(y^{*t-1}, y^{t-1}, \varepsilon_t^*;\theta) \\ y_t & = r_t(y^{*t}, y^{t-1}, \varepsilon_t;\theta) \end{array} \right. \tag{11}$$

where $\{\varepsilon_t^*\}$ and $\{\varepsilon_t\}$ are independent white noises, the marginal distributions of which may depend on θ but are absolutely continuous with respect to the same fixed measures m_* and m, 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 $r_t]$. In this case we can choose μ_{*t} as the image of m_* by $r_t^*(y^{*t-1}, y^{t-1}, \cdot;\theta)$ and μ_t as the image of m by $r_t(y^{*t}, y^{t-1}, \cdot;\theta)$.
Let us consider some examples, some of which have not been satisfactorily treated in the literature so far.

2.2.1 Factor models

A one factor model is defined by:

$$\left\{ \begin{array}{ll} y_t^* & = r_t^*(y^{*t-1}, y^{t-1}, \varepsilon_t^*;\theta), \\ y_t & = a(\theta) y_t^* + \varepsilon_t, \end{array} \right. \tag{12}$$

where $\{\varepsilon_t^*\}$ and $\{\varepsilon_t\}$ are independent white noises, with dimensions 1 and p respectively. The distributions of ε_t^* and ε_t are assumed to have p.d.f. with respect to λ_1, the Lebesgue measure on $\mathbb{R}$, and λ_p the Lebesgue measure on $\mathbb{R}^p$ and may depend on θ; $a(\theta)$ is a p-vector of coefficients.
If the function $r_t^*(y^{*t-1}, y^{t-1}, \cdot; \theta)$ is one-to-one and differentiable, y_t^* also has a p.d.f. with respect to λ_1, denoted by $f_t(y_t^*|\mathcal{I}_{t-1}; \theta)$. In this case we can take:

$$\begin{aligned}
&\mu_{*t} = m_* = \lambda_1, \\
&\mu_t = m = \lambda_p, \\
&\nu_t = \mu_{*t} \otimes \mu_t = \lambda_{p+1}.
\end{aligned}$$

Denoting by $\psi(\varepsilon; \theta)$ the p.d.f. of ε_t with respect to λ_p, the various p.d.f. considered in the general presentation are:

$$\begin{array}{ll}
f_t(y_t^*|\mathcal{I}_{t-1};\theta) & \\
f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta) & = \psi[y_t - a(\theta)y_t^*;\theta] \\
f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta) & = f_t(y_t^*|\mathcal{I}_{t-1};\theta)\psi[y_t - a(\theta)y_t^*;\theta] \\
f_t(y_t|\mathcal{I}_{t-1};\theta) & = \displaystyle\int f_t(y_t^*|\mathcal{I}_{t-1};\theta)\psi[y_t - a(\theta)y_t^*;\theta]d\lambda_1(y_t^*) \\
f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) & = \dfrac{f_t(y_t^*|\mathcal{I}_{t-1};\theta)\psi[y_t - a(\theta)y_t^*;\theta]}{\displaystyle\int f_t(y_t^*|\mathcal{I}_{t-1};\theta)\psi[y_t - a(\theta)y_t^*;\theta]d\lambda_1(y_t^*)}.
\end{array} \tag{13}$$

2.2.2 Stochastic volatility models

These models are defined by:

$$\begin{cases} y_t^* & = m_t(y^{*t-1}, y^{t-1}; \theta) + \sigma_t(y^{*t-1}, y^{t-1}; \theta)\varepsilon_t^*, \\ y_t & = \exp(0.5\ y_t^*)\varepsilon_t, \end{cases} \tag{14}$$

where $\{\varepsilon_t^*\}$ and $\{\varepsilon_t\}$ are independent scalar white noises (see Taylor [48], Ghysels, Harvey and Renault [19], Shephard [46]).
The remarks of the previous section are immediately transposed to this kind of models; we only have to change p into 1 and $\psi[y_t - a(\theta)y_t^*; \theta]$ into $\exp(-0.5\ y_t^*)\psi[y_t \exp(-0.5\ y_t^*); \theta]$.
This kind of models can be extended to stochastic volatility models with leverage effect (see Section 3.6).

2.2.3 Switching regime dynamic models

Such models are defined as:

$$\begin{cases} y_t^*|y_{t-1}^* & \sim \pi_{y_{t-1}^* y_t^*}, \\ \qquad y_t & = r_t(y^{*t}, y^{t-1}, \varepsilon_t; \theta), \end{cases} \tag{15}$$

where y_t^* is a K-state Markov chain with transition probability $\pi_{..}$, which may depend on y^{t-1} and θ, $\{\varepsilon_t\}$ is a p-dimensional white noise, which has a p.d.f. with respect to λ_p (see Kim [28], Billio and Monfort [3], Lee [32]). The function $r_t(y^{*t}, y^{t-1}, \cdot; \theta)$ is one-to-one and differentiable, which implies that the conditional distribution of y_t given (y^{*t}, y^{t-1}) has a p.d.f. with respect to λ_p, denoted by $f_t(y_t|\mathcal{I}_{t-1}, y_t^*; \theta)$.
In particular with two regimes (K=2), y_t^* can simply be described as follows:

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

where $\{\varepsilon_t^*\}$ is a white noise with distribution $\mathcal{U}_{[0,1]}$, the uniform distribution on [0,1], independent of $\{\varepsilon_t\}$. In this case y_t^* is a binary variable taking the values 0 and 1, corresponding to the two regimes, and we can choose:

$$\begin{aligned}
m_* &= \lambda_1, \\
\mu_{*t} &= \delta_0 + \delta_1, \\
m &= \mu_t = \lambda_p, \\
\nu_t &= (\delta_0 + \delta_1) \otimes \lambda_p,
\end{aligned}$$

where δ_x is the Dirac point mass at x. The various p.d.f.'s of interest are:

$$\begin{aligned}
&f_t(y_t|\mathcal{I}_{t-1}, y_t^*; \theta) \\
f_t(y_t^*|\mathcal{I}_{t-1}; \theta) = \;& \mathbb{1}_{\{0\}}(y_{t-1}^*)[\pi_{00}\mathbb{1}_{\{0\}}(y_t^*) + (1-\pi_{00})\mathbb{1}_{\{1\}}(y_t^*)] \\
&+\mathbb{1}_{\{1\}}(y_{t-1}^*)[(1-\pi_{11})\mathbb{1}_{\{0\}}(y_t^*) + \pi_{11}\mathbb{1}_{\{1\}}(y_t^*)] \\
f_t(y_t^*, y_t|\mathcal{I}_{t-1}; \theta) = \;& f_t(y_t|\mathcal{I}_{t-1}, y_t^*; \theta) f_t(y_t^*|\mathcal{I}_{t-1}; \theta) \\
f_t(y_t|\mathcal{I}_{t-1}; \theta) = \;& f_t(y_t|\mathcal{I}_{t-1}, 0; \theta)[\pi_{00}\mathbb{1}_{\{0\}}(y_{t-1}^*) + (1-\pi_{11})\mathbb{1}_{\{1\}}(y_{t-1}^*)] \\
&+f_t(y_t|\mathcal{I}_{t-1}, 1; \theta)[(1-\pi_{00})\mathbb{1}_{\{0\}}(y_{t-1}^*) + \pi_{11}\mathbb{1}_{\{1\}}(y_{t-1}^*)] \\
f_t(y_t^*|\mathcal{I}_{t-1}, y_t; \theta) = \;& \frac{f_t(y_t^*, y_t|\mathcal{I}_{t-1}; \theta)}{f_t(y_t|\mathcal{I}_{t-1}; \theta)}.
\end{aligned} \tag{16}$$

2.3 The "functional" latent models

In this case we assume that:

$$y_t = g_t(y_t^*), \tag{17}$$

where g_t is a known function. Thus, y_t^* is a deterministic transform of the latent variable, generally not one-to-one. This implies that the conditional distribution of y_t given $(\mathcal{I}_{t-1}, y_t^*)$ is the point mass at $g_t(y_t^*)$ and, therefore, we can choose:

$$\begin{aligned}
\mu_t^{y^*} &= \delta_{g_t(y^*)}, \\
\nu_t &= \mu_{*t} \otimes \delta_{g_t(\cdot)},
\end{aligned}$$

if μ_{*t} is a dominating measure for the conditional distribution of y_t^* given $\mathcal{I}_{t-1}$,

$$\mu_t = \mu_{*t} \circ g_t^{-1},$$

the image of μ_{*t} by g_t, and μ_{*t}^y as the quasi unique set of measures satisfying:

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

Note that the support of μ_{*t}^y is $g_t^{-1}(y)$.
Let us consider some examples, some of which are still under intensive study.

2.3.1 Dynamic Probit or Tobit models

We assume that the conditional distribution of y_t^* given $\mathcal{I}_{t-1} = \{y^{*t-1}\}$ is absolutely continuous with respect to λ_1. The conditional p.d.f. is denoted by $\tilde{f}_t(y_t^*|y^{*t-1}; \theta)$: it may have an explicit form, for instance if y_t^* follows a Gaussian AR(p) model, or it may be given by the Kalman filter if y_t^* follows a Gaussian linear state space model. The observable y_t is given by:

$$\begin{aligned}
y_t &= \mathbb{1}_{I\!R^+}(y_t^*) && \text{(Probit case)}, \\
y_t &= \max(y_t^*, 0) && \text{(Tobit case)}.
\end{aligned}$$

(see Maddala [36], Manrique and Shephard [37]). Let us, for instance, consider the Tobit case. We can take:

$$\begin{aligned}
\mu_{*t} &= \lambda_1, \\
\mu_t &= \delta_0 + \lambda_{I\!R^+}, \\
\mu_t^{y^*} &= \delta_{\max(y^*,0)}, \\
\mu_{*t}^{y} &= 1\!\!1_{\{0\}}(y)\lambda_{I\!R^-} + 1\!\!1_{\{I\!R^+-0\}}(y)\delta_y.
\end{aligned}$$

The p.d.f.'s are:

$$\begin{aligned}
f_t(y_t^*, y_t|\mathcal{I}_{t-1};\theta) &= \tilde{f}_t(y_t^*|y^{*t-1};\theta) \\
f_t(y_t^*|\mathcal{I}_{t-1};\theta) &= \tilde{f}_t(y_t^*|y^{*t-1};\theta) \\
f_t(y_t|\mathcal{I}_{t-1};\theta) &= F_{0t}(y^{*t-1};\theta)1\!\!1_{\{0\}}(y_t) + \tilde{f}_t(y_t|y^{*t-1};\theta)1\!\!1_{\{I\!R^+-0\}}(y_t),
\end{aligned} \tag{18}$$

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

$$\begin{aligned}
f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta) &= 1 \\
f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) &= \frac{\tilde{f}_t(y_t^*|y^{*t-1};\theta)}{F_{0t}(y^{*t-1};\theta)}1\!\!1_{\{0\}}(y_t) + 1\!\!1_{\{I\!R^+-0\}}(y_t).
\end{aligned}$$

The conditional distribution of y_t^* given $(\mathcal{I}_{t-1}, y_t)$ is:

$$[f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta)\cdot\mu_{*t}^{y_t}] = \frac{1\!\!1_{\{0\}}(y_t)1\!\!1_{\{I\!R^-\}}(y_t^*)\tilde{f}_t(y_t^*|y^{*t-1};\theta)}{F_{0t}(y^{*t-1};\theta)}\lambda_1 + 1\!\!1_{\{I\!R^+-0\}}(y_t)\delta_{y_t}. \tag{19}$$

2.3.2 Dynamic disequilibrium models

Consider $y_t^* = (y_{1t}^*, y_{2t}^*)$, which follows some bivariate dynamic model. We assume that the conditional distribution of y_t^* given $\mathcal{I}_{t-1} = \{y^{*t-1}\}$ has a density function $\tilde{f}_t(y_t^*|y^{*t-1};\theta)$ with respect to λ_2, the Lesbegue measure on $I\!R^2$. The observable y_t is given by:

$$y_t = \min(y_{1t}^*, y_{2t}^*).$$

(see Laroque and Salanié [31], Lee [32], Hajivassiliou and McFadden [22]). Here we can take:

$$\begin{aligned}
\mu_{*t} &= \lambda_2, \\
\mu_t &= \lambda_1, \\
\mu_t^{y^*} &= \delta_{\min(y_1^*,y_2^*)}, \\
\mu_{*t}^{y} &= \delta_y \otimes \lambda_{[y,\infty[} + \lambda_{[y,\infty[} \otimes \delta_y.
\end{aligned}$$

The p.d.f.'s are:

$$\begin{aligned}
f_t(y_{1t}^*, y_{2t}^*, y_t|\mathcal{I}_{t-1};\theta) &= \tilde{f}_t(y_{1t}^*, y_{2t}^*|y^{*t-1};\theta) \\
f_t(y_{1t}^*, y_{2t}^*|\mathcal{I}_{t-1};\theta) &= \tilde{f}_t(y_{1t}^*, y_{2t}^*|y^{*t-1};\theta) \\
f_t(y_t|\mathcal{I}_{t-1};\theta) &= \int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|\mathcal{I}_{t-1};\theta)d\lambda_1(y_2^*) + \int_{[y_t,\infty[} \tilde{f}_t(y_1^*, y_t|\mathcal{I}_{t-1};\theta)d\lambda_1(y_1^*),
\end{aligned} \tag{20}$$

and therefore:

$$\begin{aligned}
f_t(y_t|\mathcal{I}_{t-1}, y_t^*;\theta) &= 1 \\
f_t(y_t^*|\mathcal{I}_{t-1}, y_t;\theta) &= \frac{\tilde{f}_t(y_t^*|y^{*t-1};\theta)}{\displaystyle\int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|\mathcal{I}_{t-1};\theta)d\lambda_1(y_2^*) + \int_{[y_t,\infty[} \tilde{f}_t(y_1^*, y_t|\mathcal{I}_{t-1};\theta)d\lambda_1(y_1^*)}.
\end{aligned}$$

The conditional distribution of y_t^* given $(\mathcal{I}_{t-1}, y_t)$ is:

$$[f_t(y_t^*|\mathcal{I}_{t-1}, y_t; \theta) \cdot \mu_{*t}^{y_t}] = \frac{\delta_{y_t} \otimes [\tilde{f}_t(y_t, y_{2t}^*|y^{*t-1}; \theta) \cdot \lambda_{[y_t,\infty[}] + [\tilde{f}_t(y_{1t}^*, y_t|y^{*t-1}; \theta) \cdot \lambda_{[y_t,\infty[}] \otimes \delta_{y_t}}{\int_{[y_t,\infty[} \tilde{f}_t(y_t, y_2^*|\mathcal{I}_{t-1}; \theta) d\lambda_1(y_2^*) + \int_{[y_t,\infty[} \tilde{f}_t(y_1^*, y_t|\mathcal{I}_{t-1}; \theta) d\lambda_1(y_1^*)}. \tag{21}$$

2.3.3 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 dates.
This continuous time process can be approximated as closely as desired by the Euler method. Using, for this approximation, a time unit which is n times smaller than the lag between two observations, we can write the model:

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

where ε_t is a standard Gaussian white noise and $\mathbb{1}_{\{n\mathbb{N}\}}(t)$ is equal to 1 if t is a multiple of n and to 0 otherwise.
This setup simply says that y_t^* is observed at the K dates 1,n+1,2n+1,...,(K-1)n+1=T, and that y_t is equal to y_t^* at these dates and to zero (by convention) otherwise. In this model we can take:

$$\begin{aligned} \mu_{*t} &= \lambda_1, \\ \mu_t &= \begin{cases} \lambda_1 & \text{if } t-1 \in n\mathbb{N}, \\ \delta_0 & \text{otherwise} \end{cases} \\ \mu_t^{y^*} &= \begin{cases} \delta_{y^*} & \text{if } t-1 \in n\mathbb{N}, \\ \delta_0 & \text{otherwise} \end{cases} \\ \mu_{*t}^{y} &= \begin{cases} \delta_y & \text{if } t-1 \in n\mathbb{N}, \\ \lambda_1 & \text{otherwise.} \end{cases} \end{aligned}$$

The p.d.f.'s are:

$$\begin{aligned} f_t(y_t^*, y_t|\mathcal{I}_{t-1}; \theta) &= \tilde{f}(y_t^*|y_{t-1}^*; \theta) \\ &= \frac{\sqrt{n}}{b(y_{t-1}^*; \theta)} \phi \left[\frac{\sqrt{n}(y_t^* - y_{t-1}^* - a(y_{t-1}^*; \theta)/n)}{b(y_{t-1}^*; \theta)} \right] \\ f_t(y_t^*|\mathcal{I}_{t-1}; \theta) &= \tilde{f}(y_t^*|y_{t-1}^*; \theta) \\ f_t(y_t|\mathcal{I}_{t-1}; \theta) &= \begin{cases} \tilde{f}(y_t^*|y_{t-1}^*; \theta) & \text{if } t-1 \in n\mathbb{N}, \\ 1 & \text{otherwise} \end{cases} \\ f_t(y_t|\mathcal{I}_{t-1}, y_t^*; \theta) &= 1 \\ f_t(y_t^*|\mathcal{I}_{t-1}, y_t; \theta) &= \begin{cases} 1 & \text{if } t-1 \in n\mathbb{N}, \\ \tilde{f}(y_t^*|y_{t-1}^*; \theta) & \text{otherwise} \end{cases} \end{aligned} \tag{22}$$

where ϕ is the p.d.f. of N(0,1).

3 THE SIMULATED LIKELIHOOD RATIO (SLR) METHOD

3.1 The general principle

The method will be based on the following result, which expresses the likelihood ratio as the expectation of a computable ratio. The same device is used by Geyer [14] in a different setup: we generalize his result in a dynamic context with potential functional relationships, as in Section 2.3 (see also Geyer and Thomson [15]). It actually relates to the general principle of importance sampling (see Geweke [17]).

Proposition 2
In the setup of Section 2, and for any values θ and $\bar{\theta}$ of the parameters, and any integer $k \in \{1, ..., T\}$, the following result holds:

$$\frac{f(y^T;\theta)}{f(y^T;\bar{\theta})} = E_{\bar{\theta}}\left[\frac{f(y^{*T-k}, y^T;\theta)}{f(y^{*T-k}, y^T;\bar{\theta})}\middle| y^T\right], \tag{23}$$

where $f(y^{*T-k}, y^T;\theta)$ is the p.d.f. of (y^{*T-k}, y^T) with respect to $\otimes_{t=1}^{T-k}\nu_t \otimes_{t=T-k+1}^{T}\mu_t$.

Proof: see Appendix 2.
As detailed below, the previous result is particularly useful for the case $k = 0$. However the case $k > 0$ is also interesting when dealing with importance sampling of order greater than 1 (see Section 3.2).
It is clear that the maximization of the likelihood function $f(y^T;\theta)$ is equivalent to the maximization of the ratio $\frac{f(y^T;\theta)}{f(y^T;\bar{\theta})}$, where $\bar{\theta}$ is a given value of the parameter. Since, in the models considered, $f(y^T;\theta)$ is not computable, the same is true for this ratio. However the previous result shows that it can be expressed as an expectation and, therefore, it can be, in principle, approximated by simulations.
Let us denote by $y^{*T-k}(s)$, $s = 1, ..., S$, simulated paths in the conditional distribution of y^{*T-k} given y^T associated with the value $\bar{\theta}$ of the parameter, i.e. from $f(y^{*T-k}|y^T, \bar{\theta})$.
The Simulated Likelihood Ratio (SLR) method will amount to maximizing:

$$\frac{1}{S}\sum_{s=1}^{S}\frac{f(y^{*T-k}(s), y^T;\theta)}{f(y^{*T-k}(s), y^T;\bar{\theta})} \tag{24}$$

with respect to θ.
The problem which must be solved now is the simulation in this conditional distribution, and this problem will be dealt with in the following sections. But let us discuss before some attractive features of this method.

First, provided that S is large enough, the SLR method can be considered as asymptotically efficient, which is not the case of Methods of Simulated Moments, Simulated Pseudo Likelihood Methods or Indirect Inference Methods.

Second, and contrary to Simulated EM methods, such as SEM (Celeux and Diebolt [5], Diebolt and Ip [8]) and MCEM (Wei and Tanner [50]), it necessitates only one optimization run (and not a sequence of optimizations) and it allows for the computation of the likelihood surface and likelihood ratio test statistics.

Third, it is important to note that the simulation is made for only one value of the parameter and, therefore, the objective function will be smooth with respect to θ, even if the simulation involves rejection methods or if the endogenous variables are non differentiable, or discontinuous function of the parameter. Moreover, in some situations, when the simulation is time consuming, it may be interesting to store the simulated paths, instead of redrawing them (with the same seed), when computing (23) for different values of θ. This is the case when convergence of (24) 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 [14], 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 following $\bar{\theta}$.

Fourth, since each term of (23) 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 avoided. A related issue is variance finiteness: when

$$E_{\bar{\theta}}\left[\left(\frac{f(y^{*T-k}, y^T; \theta)}{f(y^{*T-k}, y^T; \bar{\theta})}\right)^2 \middle| y^T\right] < \infty \tag{25}$$

does not hold, (24) has an infinite variance and may converges to (23) at a slower rate. However, this problem did not occur in the applications presented below, in the sense that we did not observe sudden changes in the likelihood approximation in certain regions of the parameter space.

3.2 Implementation

As mentioned above, the method can be implemented by simulating in 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. $f(y^{*T-k}|y^T;\bar{\theta})$ with respect to the product measure $\otimes_{t=1}^{T-k}\mu_{*t}^{y_t}$. Since it is, in general, impossible to simulate directly in this distribution, a natural idea is to try to use a Gibbs sampling approach, i.e. to draw successively in the conditional distribution $[f_t(y_t^*|y_{-t}^*, y^T;\bar{\theta}) \cdot \mu_{*t}^{y_t}]$, with $y_{-t}^* = (y_1^*, ..., y_{t-1}^*, y_{t+1}^*, ..., y_{T-k}^*)$. However, generating such $y_t^*(s)$'s may also prove difficult to achieve and may be replaced with an Metropolis-Hastings step, i.e. by generating $z_t(s)$ from an instrumental (or candidate) p.d.f. $q_t(z_t|z_{t-1}(s), y^{*T-k}(s\text{-}1), y^T)$. Then, at iteration s of the Gibbs sampler, the whole simulated path $y^{*T-k}(s)$ is taken 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:

$$\rho = \min\left\{1, \frac{f(z^{T-k}(s)|y^T;\bar{\theta})\prod_{t=1}^{T-k} q_t(y_t^*(s\text{-}1)|y^{*t-1}(s\text{-}1), z^{T-k}(s), y^T)}{f(y^{*T-k}(s\text{-}1)|y^T;\bar{\theta})\prod_{t=1}^{T-k} q_t(z_t(s)|z^{t-1}(s), y^{*T-k}(s\text{-}1), y^T)}\right\}. \tag{26}$$

Note that, although $f(y^{*T-k}|y^T;\bar{\theta})$ is not computable, the ratio in ρ is computable 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 computable. This global strategy in which the whole sequence $z^{T-k}(s)$ is globally accepted or rejected may lead to low acceptance rates, in cases of poor fit between $f(z^{T-k}|y^T;\bar{\theta})$ and the proposal and may then be replaced with a multimove strategy in which the sequence are partitioned into blocks, as in Liu, Wong and Kong [35] or Shephard and Pitt[4] [47]. At iteration s, a block $z_p^r(s) = (z_p(s),...,z_r(s))$, with $p \leq r$, 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:

$$\rho = \min \left\{ 1, \frac{f(z_p^r(s)|y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1), y^T)}{f(y_p^{*r}(s\text{-}1)|y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1), y^T)} \right. \\ \left. \times \prod_{t=p}^{r} \frac{q_t(y_t^*(s\text{-}1)|y_p^{*t-1}(s\text{-}1), z_p^r(s), y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1))}{q_t(z_t(s)|z_p^{t-1}(s), y_p^{*r}(s\text{-}1), y_1^{*p-1}(s), y_{r+1}^{*T-k}(s\text{-}1))} \right\}. \tag{27}$$

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 formulae (7) and the terms of such a factorization in which no y_t^* variable at dates $t = p,...,r$ appears cancel out.
The extreme version of this multimove strategy is the "one move" case in which $r = p$. In this case the first set of conditioning variables in q_t is always empty.
An important issue for the performance of the method pertains to the choice of the instrumental distribution q_t. We consider different aspects of this choice in the following paragraphs.

3.3 Instrumental distribution of order 1

We take $k = 0$ in (23) and:

$$q_t(y_t^*|y^{*t-1}, y^{*T}(s\text{-}1)) = f_t(y_t^*|y^{*t-1}, y^t;\bar{\theta}), \tag{28}$$

which is an importance sampling p.d.f., sometimes called the importance sampling function of order 1 (see Billio and Monfort [3]). Note that q_t does not depend on

[4]The theoretical gain resulting from grouping must be, however, backed by a good choice of the candidate distributions.

$y^{*T}(s\text{-}1)$ and therefore leads to an independent Metropolis-Hastings approach (Tierney [49]).
In this case, and if we adopt a global strategy, the acceptance probability ρ is equal to:

$$\begin{aligned}
\rho &= \min\left\{1, \frac{f(z^T(s)|y^T;\bar\theta)\prod_{t=1}^{T} f_t(y_t^*(s\text{-}1)|y^{*t-1}(s\text{-}1),y^t;\bar\theta)}{f(y^{*T}(s\text{-}1)|y^T;\bar\theta)\prod_{t=1}^{T} f_t(z_t(s)|z^{t-1}(s),y^t;\bar\theta)}\right\} \\
&= \min\left\{1, \frac{\prod_{t=1}^{T} f_t(y_t|z^{t-1}(s),y^{t-1};\bar\theta)}{\prod_{t=1}^{T} f_t(y_t|y^{*t-1}(s\text{-}1),y^{t-1};\bar\theta)}\right\}.
\end{aligned} \tag{29}$$

In some settings, both the simulation from (28) and the computation of (29) are manageable: this is the case for some factor models, switching regime dynamic models, the dynamic Probit - Tobit models, the dynamic disequilibrium models and continuous time models. Let us consider these models.

Example 3.1 (Factor ARCH models). Consider, for instance, a Gaussian factor ARCH model defined by:

$$\begin{cases} y_t^* &= (\alpha+\beta y_{t-1}^{*2})^{1/2}\varepsilon_t^*, \\ y_t &= a y_t^* + \varepsilon_t, \end{cases} \tag{30}$$

where $\varepsilon_t^* \overset{iid}{\sim} N(0,1)$, $\varepsilon_t \overset{iid}{\sim} N(0,\Sigma)$ 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 $\bar\theta = (\bar\alpha,\bar\beta,\bar a,\bar\Sigma)$,

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

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

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

where ϕ is the p.d.f. of $N(0,1)$. 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 3.2 (Switching regime dynamic models). In the switching regime dynamic models of Section 2.2.3, 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 formulae (16)). It is the case in the general framework of Switching State Space Models (Billio and Monfort [3]), where these p.d.f.'s are obtained through the Partial Kalman Filter.
It is also the case in a switching GARCH process like:

$$\begin{cases} y_t^* &= \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= \alpha + \beta y_t^* + \gamma y_{t-1}^2 + \delta \sigma_{t-1}^2, \end{cases} \tag{31}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{U}_{[0,1]}$ and $\varepsilon_t \overset{iid}{\sim} N(0,1)$ independently, $\theta = (\alpha, \beta, \gamma, \delta, \pi_{00}, \pi_{11})$. Here $f_t(y_t|y^{*t}, y^{t-1}; \bar{\theta})$ is the Gaussian p.d.f. that would be obtained for the GARCH process if y_t^*'s were deterministic, i.e. the Gaussian p.d.f. with zero mean and a 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 then easily done. Similarly, the p.d.f.'s $f_t(y_t|y^{*t-1}, y^{t-1}; \bar{\theta})$ are available (see formulae (16)), and, therefore, the acceptance probability ρ, given in (29), is also easily obtained.

Example 3.3 (Dynamic Tobit models). In the dynamic Tobit case of Section 2.3.1 the conditional distribution of y_t^* given (y^{*t-1}, y^t) is given by (19). The simulation in this distribution is easily done:

$$\begin{aligned} z_t(s) &= y_t, && \text{if } y_t > 0, \\ z_t(s) &\sim \frac{f_t(z_t|z^{t-1}(s); \bar{\theta})}{F_{0t}(z^{t-1}(s); \bar{\theta})} \mathbb{1}_{\mathbb{R}^-}(z_t), && \text{if } y_t = 0. \end{aligned}$$

The second case is simply simulating from a truncated distribution. We know that this can be done smoothly in the normal case but, as noted in Section 3.1, the SLR method does not require smoothness and, in non-normal cases, rejection methods can be used[5].
The p.d.f.'s $f_t(y_t|y^{*t-1}, y^{t-1}; \bar{\theta})$ are simply (see (18)):

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

Therefore ρ is easily computed.

Example 3.4 (Dynamic disequilibrium models). In the dynamic disequilibrium model of Section 2.3.2, the conditional distribution of y_t^* given (y^{*t-1}, y^t) is given by (21). The simulation in this distribution is straightforward:

- first draw the regimes:

  1: with probability

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

[5] See Geweke [18] or Robert [41] for optimal generation algorithms in the truncated normal case, and Philippe [40] for the corresponding truncated Gamma case.

2: with probability $1-\pi$,

- then draw $z_t(s)$ in the following way:

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 Gaussian case the previous drawings are simply done in truncated Gaussian distributions. The p.d.f.'s $f_t(y_t|y^{*t-1}, y^{t-1};\bar{\theta})$ are given in (20) and, therefore, the acceptance probability ρ is also easily computed.

Example 3.5 (Continuous time models). In the case of a continuous time model as in Section 2.3.3, the conditional distribution of y_t^* given (y^{*t-1}, y^t) is δ_{y_t} if t-1$\in n\mathbb{N}$, $\tilde{f}(y_t^*|y_{t-1}^*;\bar{\theta})$ otherwise (see formulae (22)). The acceptance probability is then:

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

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:

$$\begin{array}{ll} \tilde{f}(z_{(k-1)n+1}|y^*_{(k-1)n}(s);\bar{\theta}), & \\ \tilde{f}(z_t|z_{t-1}(s);\bar{\theta}), & t=(k-1)n+2,...,kn. \end{array}$$

The acceptance probability is:

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

with

$$\begin{aligned} A[z^{kn}_{(k-1)n+1}(s)] &= \frac{\tilde{f}(z_{(k-1)n+1}(s)|y^*_{(k-1)n}(s);\bar{\theta})}{\tilde{f}(z_{(k-1)n+1}(s)|y^*_{(k-1)n}(s);\bar{\theta})} \\ &\times \frac{\prod_{t=(k-1)n+2}^{kn} \tilde{f}(z_t(s)|z_{t-1}(s);\bar{\theta})\tilde{f}(y^*_{kn+1}(s\text{-}1)|z_{kn}(s);\bar{\theta})}{\prod_{t=(k-1)n+2}^{kn} \tilde{f}(z_t(s)|z_{t-1}(s);\bar{\theta})} \\ &= \tilde{f}(y_{kn+1}|z_{kn}(s);\bar{\theta}). \end{aligned}$$

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 Instrumental distribution of order 2

We also consider another case of interest, when $k = 1$ in (23) 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

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

(see Appendix 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:

$$\rho = \min\left\{1, \frac{f(z^{T-1}(s)|y^T;\bar{\theta})\prod_{t=1}^{T-1} f_t(y_t^*(s\text{-}1)|y^{*t-1}(s\text{-}1), y^{t+1};\bar{\theta})}{f(y^{*T-1}(s\text{-}1)|y^T;\bar{\theta})\prod_{t=1}^{T-1} f_t(z_t(s)|z^{t-1}(s), y^{t+1};\bar{\theta})}\right\}. \tag{32}$$

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 can be applied if it is easy to simulate in $f_t(y_t^*|y^{*t-1}, y^{t+1};\bar{\theta})$ and if $f_{t+1}(y_{t+1}|y^{*t-1}, y^t;\bar{\theta})$ is easily computable; it is generally the case if y_t^* takes a finite number of values (see Billio and Monfort [3] for the Switching State Space case).

3.5 Instrumental distribution with future values of y_t^*

We take $k = 0$ in (23) and:

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

More precisely we get:

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

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 will be easily done for instance if y^*_t can take a finite number of values, which is the case in the switching regimes models. For instance if y^*_t can take the values 0 and 1, we have $\mu^{y_t}_{*t} = \delta_0 + \delta_1$ and:

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

The acceptance probability ρ is given by (26). Since

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

it turns out that ρ becomes:

$$\rho = \min\{1, \rho^*\}, \tag{33}$$

with

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

(see Billio, Monfort and Robert [4] for an application of such an Metropolis-Hastings approach in the Bayesian framework). Note that this approach is no longer an independent approach. It could be used for the Switching State Space models and the switching GARCH models considered above (see Example 3.2).

3.6 Use of approximations

The instrumental p.d.f. can also be based on some approximations of $f_t(y^*_t|y^*_{-t}, y^T;\bar{\theta})$.

Example 3.6 (Stochastic volatility models). Let us consider a stochastic volatility model (see (14)):

$$\begin{cases} y_t^* &= m_t(y^{*t-1}, y^{t-1}; \theta) + \sigma_t(y^{*t-1}, y^{t-1}; \theta)\varepsilon_t^*, \\ y_t &= \exp(0.5\, y_t^*)\varepsilon_t, \end{cases} \tag{34}$$

where $(\varepsilon_t^*, \varepsilon_t)$ is a standard (zero mean, identity matrix) bivariate Gaussian white noise. The more classical stochastic volatility models are the basic one:

$$\begin{cases} y_t^* &= a + by_{t-1}^* + c\varepsilon_t^*, \\ y_t &= \exp(0.5\, y_t^*)\varepsilon_t, \end{cases} \tag{35}$$

and the "leverage effect" one:

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

where $(\tilde{\varepsilon}_t, \varepsilon_{t-1})$ is a Gaussian 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 in general $\rho < 0$. This model can be put in the form (34) by considering the regression of $\tilde{\varepsilon}_t$ on ε_{t-1}:

$$\tilde{\varepsilon}_t = \rho\varepsilon_{t-1} + \sqrt{1-\rho^2}\varepsilon_t^*,$$

where ε_t^* is independent of ε_{t-1}:

$$\begin{cases} y_t^* &= a + by_{t-1}^* + c\rho y_{t-1}\exp(-0.5\, y_{t-1}^*) + c\sqrt{1-\rho^2}\varepsilon_t^*, \\ y_t &= \exp(0.5\, y_t^*)\varepsilon_t. \end{cases} \tag{36}$$

For sake of simplicity, let us consider the basic model (35). The simulation of the y_t^*'s can be organized according to several approximations of the following p.d.f.

$$\begin{aligned} f(y_t^*|y_{t-1}^*, y_{t+1}^*, y^t, \theta) \quad \propto \quad & \exp\left\{-(y_t^* - a - by_{t-1}^*)^2/2c^2 \right. \\ & \left. -(y_{t+1}^* - a - by_t^*)^2/2c^2 - y_t^*/2 - y_t^2 e^{-y_t^*}/2\right\}. \end{aligned} \tag{37}$$

First, the part $(y_t^* + y_t^2 e^{-y_t^*})/2$ in the exponential can be linearized as in Shephard [44], Shephard and Pitt [47]. That is, it can be approximated by $(y_t^* - \log(y_t^2))^2/4$ by a second order series expansion. The proposal distribution is then a Gaussian distribution

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

with the appropriate Metropolis-Hastings acceptance probability.
A second solution follows from Jacquier, Polson and Rossi [27] and it is based on a Gamma approximation for $\omega_t = \exp(-y_t^*)$. In fact, the distribution on ω_t corresponding to (37) is given by

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

We can then approximate the second part of this distribution by a Gamma distribution $\mathcal{G}(\alpha, \beta)$ on ω_t, that is $\omega_t \sim \pi(\omega_t) \propto \omega_t^\alpha \exp(-\beta\omega_t)$, by either matching the two first moments, as in Jacquier, Polson and Rossi [27], 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 [23], [24] and many others have used autoregressions with Markov regime switching to model 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 [30], who deals with a very special case for which it is possible to apply the Hamilton's filter, Kim [28], Billio and Monfort [3] and Lee [32]. Kim [28] suggests a recursive approximation of the likelihood while Billio and Monfort [3] and Lee [32] propose a simulated maximum likelihood approach.

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

$$\begin{cases} y_t^* &= \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \mu_0 + (\mu_1 - \mu_0)y_t^* + \sigma\varepsilon_t + \alpha\sigma\varepsilon_{t-1}, \end{cases} \tag{38}$$

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

We consider the following 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 Section 3.4 and Billio, Monfort and Robert [4]):

Set $\eta_0^{(s)} = 0$, draw z_0 from the invariant distribution associated with $(\bar\pi_{00}, \bar\pi_{11})$ and, 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 $\eta_t^{(s)} = y_t - \bar\mu_{z_t} - \bar\alpha\eta_{t-1}^{(s)}$

3. 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 $\rho = \min(\omega, 1)$, with

$$\begin{aligned} \omega &= \prod_{t\geq 1} \frac{\bar\pi_{y_t^* z_{t+1}}}{\bar\pi_{z_t y_{t+1}^*}} \\ &\times \prod_{t\geq 1} \frac{\sum_{i=0}^{1} \bar\sigma^{-1}\exp\{-(y_t - \bar\mu_i - \bar\alpha\eta_{t-1}^{(s)})^2/2\bar\sigma^2\}\bar\pi_{z_{t-1}i}\bar\pi_{iy_{t+1}^*}}{\sum_{i=0}^{1} \bar\sigma^{-1}\exp\{-(y_t - \bar\mu_i - \bar\alpha\eta_{t-1}^{(s-1)})^2/2\bar\sigma^2\}\bar\pi_{y_{t-1}^* i}\bar\pi_{iz_{t+1}}}, \end{aligned}$$

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

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

We select two increasingly mixed-up models in order to evaluate the performance of the SML method through a Monte Carlo experiment. We consider 50 simulated samples of size $T = 200$ and 2000 iterations of the completion step, of which we discard the first 1000 ones. Some simulated samples are presented in Figure 1.

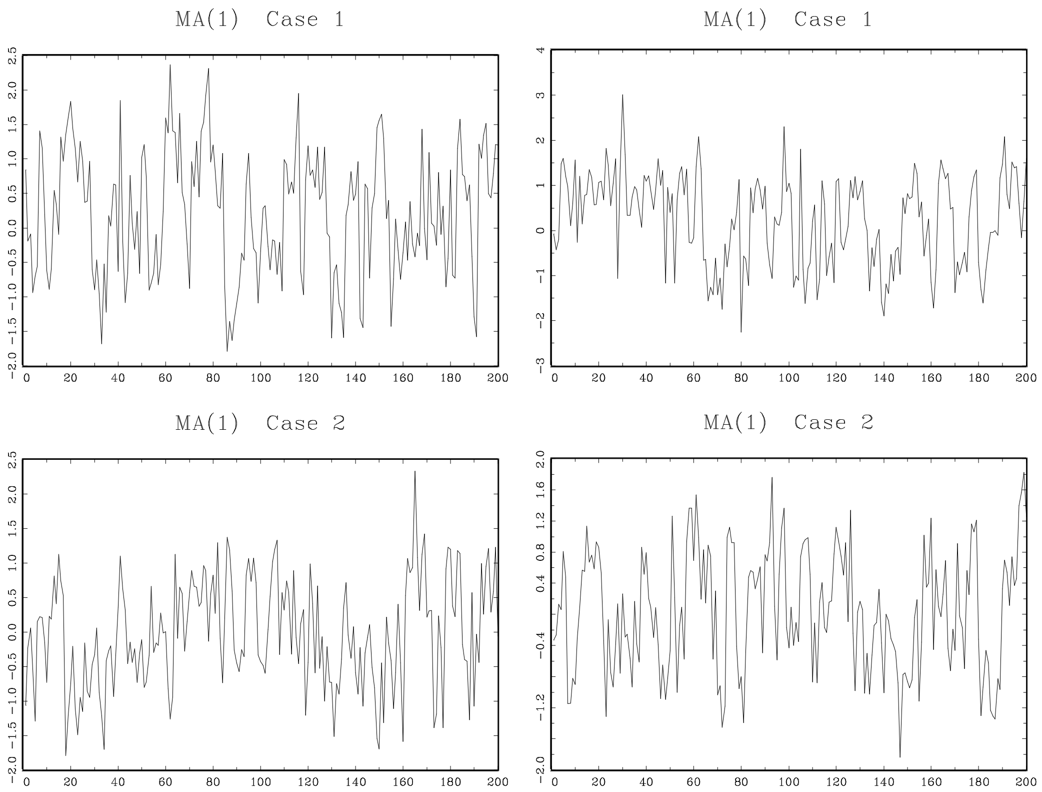

Figure 1: Simulated samples from the MA(1) model (38), case 1 and 2.

The different parameters of these models are described in Tables 1 and 2 (first column), along with the parameter values considered in the simulations ($\bar{\theta}$, 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 column, respectively). Since we are in a simulation setup, we can also compare the estimated regimes[6] versus the true ones. In both cases the correct assignment rate (adequation rate) is around 90%.
We also consider a MA(2) model with switching mean:

$$\left\{ \begin{array}{ll} y_t^* & = \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t & = \mu_0 + (\mu_1 - \mu_0)y_t^* + \sigma\varepsilon_t + \alpha\sigma\varepsilon_{t-1} + \phi\sigma\varepsilon_{t-2}, \end{array} \right. \tag{39}$$

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

[6]Given the parameter estimates, 1500 iterations of the completion step are performed and the average over the last 500 is calculated for every date t: if this average is over 0.5 we consider $y_t^* = 1$, $y_t^* = 0$ otherwise.

	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 (38) - 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 (38) - 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 (39) - MC replications 50 - acceptance rate 0.67 - adequation rate 0.93.

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 starting point of the SML algorithm has small influence, except for α and ϕ, for the success of the method. In any case, it is possible to select a good starting point such as the Bayes estimates of the parameters obtained by the algorithm proposed by Billio, Monfort and Robert [4].
Finally we consider a GARCH(1,1) model with switching constant in the variance:

$$\begin{cases} y_t^* &= \mathbb{1}_{\{0\}}(y_{t-1}^*)\mathbb{1}_{[\pi_{00},1]}(\varepsilon_t^*) + \mathbb{1}_{\{1\}}(y_{t-1}^*)\mathbb{1}_{[0,\pi_{11}]}(\varepsilon_t^*), \\ y_t &= \sigma_t\varepsilon_t, \\ \sigma_t^2 &= \alpha + \beta y_t^* + \gamma y_{t-1}^2 + \delta\sigma_{t-1}^2, \end{cases} \tag{40}$$

where $\varepsilon_t^* \overset{iid}{\sim} \mathcal{U}_{[0,1]}$ and $\varepsilon_t \overset{iid}{\sim} N(0,1)$ 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 3.2, 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 SML 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 the likelihood ratio.

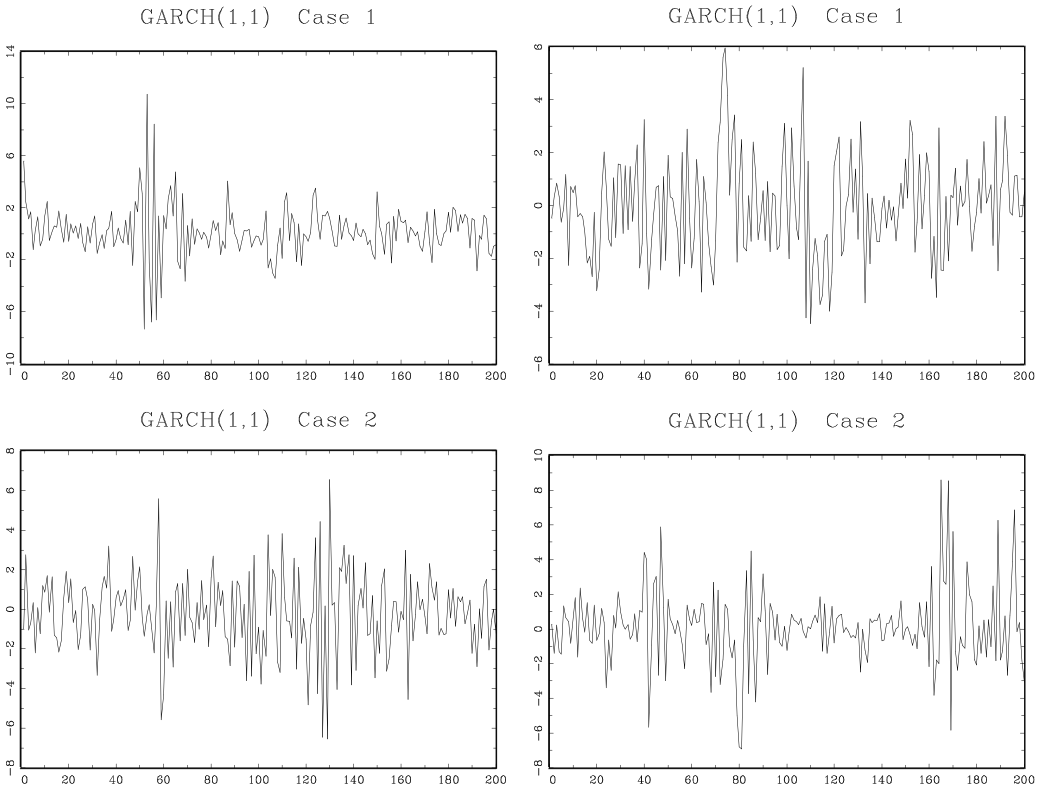

Figure 2: Simulated samples from the GARCH(1,1) model (40), case 1 and 2.

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 SML estimates are close to the true parameters and the adequation rate is around 60%.

As far as the starting point is concerned, it is important to consider reasonable values of the parameters. It is possible to obtain a valuable starting point by considering the estimates of the parameter obtained with the approximation of the likelihood suggested by Kim [28] and Dueker [9].

4.2. Dynamic Tobit models. We consider the following dynamic Tobit model:

$$\left\{ \begin{array}{ll} y_t^* & = \alpha + \beta y_{t-1}^* + \sigma \varepsilon_t^*, \\ y_t & = \max(y_t^*, 0), \end{array} \right. \tag{41}$$

	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 (40) - 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 (40) - case 2 - MC replications 50 - only two regime cases: 48 - acceptance rate 0.24 - adequation rate 0.6.

where $\varepsilon_t^* \overset{iid}{\sim} N(0,1)$. In this setting, Manrique and Shephard [37] devise an MCMC simulated EM algorithm and derive Bayes estimate.

In the simulations we use a Metropolis-Hastings algorithm, with an importance sampling p.d.f. of order 1, as described in Section 3.2 (see Example 3.3). 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 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 3.
In both cases the results are very interesting (see Tables 6 and 7): the parameter estimates are close to the true values and there is no dependence from the starting point.

	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 (41) - case 1 - MC replications 50 - acceptance rate 0.51 - zero percentage 0.4978.

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

$$\begin{cases} y_t^* &= \alpha + \beta y_{t-1}^* + \gamma y_{t-2}^* + \sigma \varepsilon_t^*, \\ y_t &= \max(y_t^*, 0). \end{cases} \tag{42}$$

We obtain the same good results, as can be seen from Table 8.

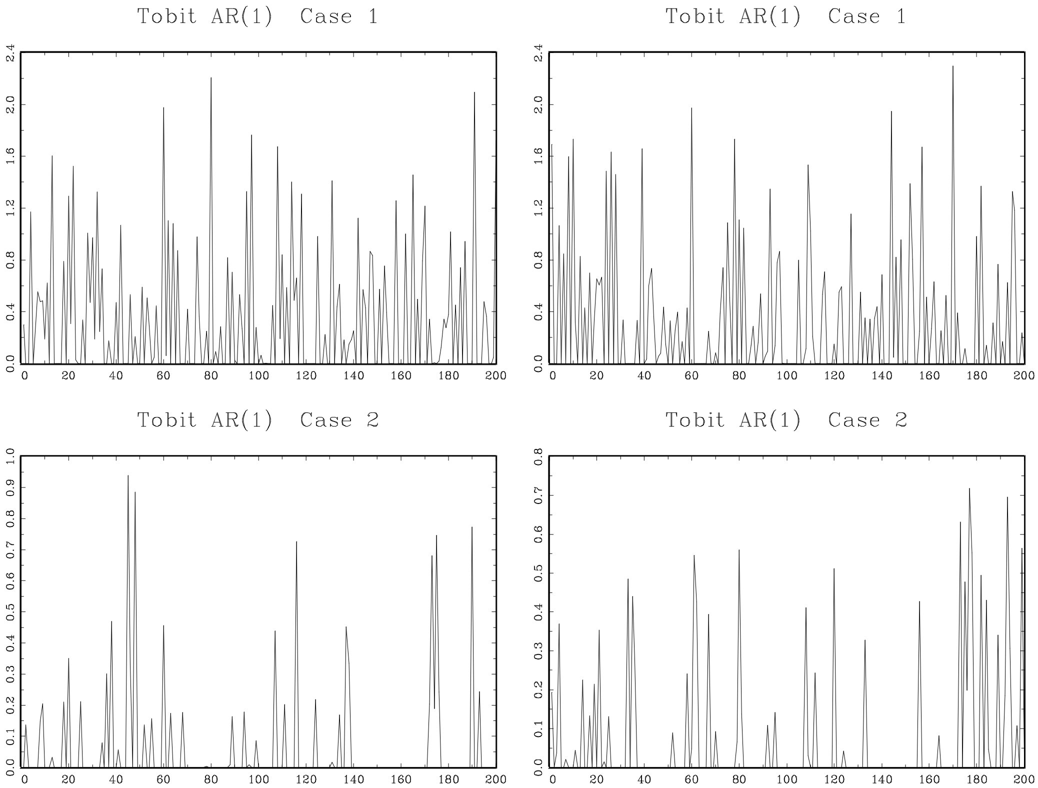

Figure 3: Simulated samples from the Tobit AR(1) model (41), case 1 and 2.

	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 (41) - 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 (42) - 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 [7], Engle, Ng and Rothschild [12]).

For simplicity and identifiability reasons, the original representation (30) is reduced to:

$$\left\{ \begin{array}{lll} y_t^* & = (1+\beta y_{t-1}^{*2})^{1/2}\varepsilon_t^* & \varepsilon_t^* \sim N(0,1), \\ y_t & = a y_t^* + \varepsilon_t & \varepsilon_t \sim N_p(0,\sigma^2 I_p), \end{array} \right. \tag{43}$$

and the dimension of y_t is $p=2$. As far as the estimation of this model is concerned Diebold and Nerlove [7] propose to apply the extended Kalman filter, which leads to some approximations, while Gouriéroux, Monfort and Renault [21] 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 sampling p.d.f. (see Section 3.3)

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

which leads to the proposal distribution

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

In order to illustrate the performances 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 Section 3.2). In the former case the acceptance probability is $\rho=\min(\omega,1)$, with

$$\begin{aligned} \omega &= \frac{\displaystyle\prod_{t=2}^{T}\frac{1}{\sqrt{(1+\bar\beta z_{t-1}^2)||\bar a||^2+\bar\sigma^2}}\exp\left\{-\frac{1}{2}y_t'[(1+\bar\beta z_{t-1}^2)\bar a\bar a' + \bar\sigma^2 I_2]^{-1}y_t\right\}}{\displaystyle\prod_{t=2}^{T}\frac{1}{\sqrt{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2+\bar\sigma^2}}\exp\left\{-\frac{1}{2}y_t'[(1+\bar\beta y_{t-1}^{*2})\bar a\bar a' + \bar\sigma^2 I_2]^{-1}y_t\right\}} \\ &= \prod_{t=2}^{T}\frac{\sqrt{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2+\bar\sigma^2}}{\sqrt{(1+\bar\beta z_{t-1}^2)||\bar a||^2+\bar\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(1+\bar\beta z_{t-1}^2)y_t'\bar a\bar a' y_t}{(1+\bar\beta z_{t-1}^2)||\bar a||^2\bar\sigma^2+\bar\sigma^4} + \frac{1}{2}\frac{(1+\bar\beta y_{t-1}^{*2})y_t'\bar a\bar a' y_t}{(1+\bar\beta y_{t-1}^{*2})||\bar a||^2\bar\sigma^2+\bar\sigma^4}\right\} \end{aligned}$$

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

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

	True	Starting value	SLR estimate
a	-0.2	-0.153	-0.14
	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 (43) with a simulated sample of size $T = 100$ and a Bayes estimate as starting value.

From the simulations, it appears that there is no difference between the two approaches in term of estimation results and then we consider the global one in the following experience.

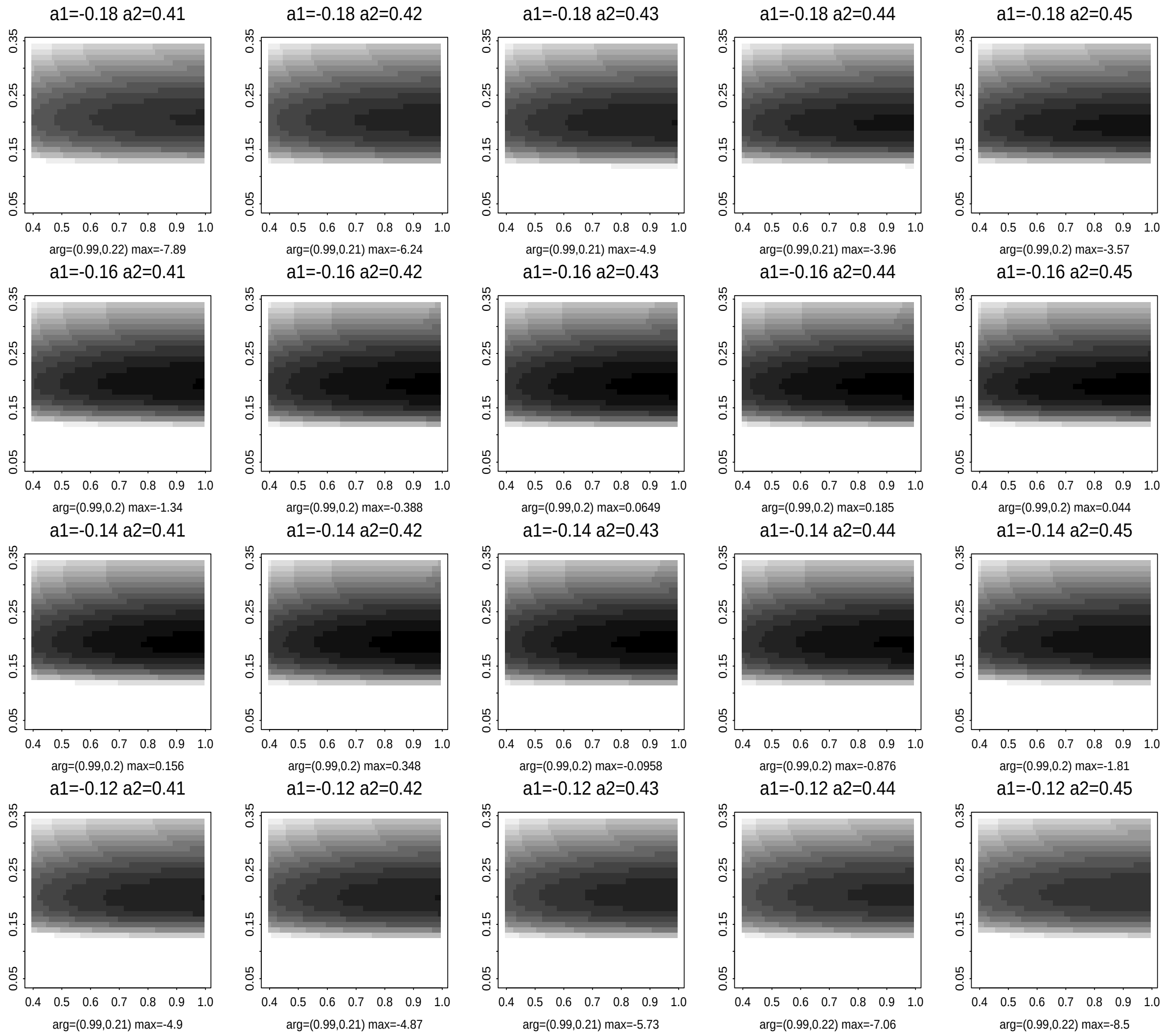

Figure 4: Estimated slices of the log-likelihood ratio for the factor ARCH model (43) for a simulated sample of size $T = 100$ and 50.000 iterations of the completion step, with a batch size of 50. See Table 9 for the parameter values.

We found that the SLR method is quite sensitive in this case to the starting value, thus we choose to initialize 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 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 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 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 [25], GMM methods of Andersen and Sorensen [1] and the indirect inference approach of Monfardini [39] are examples of the latter category. Several propositions have also been made to evaluate the likelihood function. Kim, Shephard and Chib [29] show how this function can be approximated when a mixture of normal distribution is used to describe the density of the disturbances. Jacquier, Polson and Rossi [27] consider a Bayesian approach using a Monte Carlo Markov Chain technique. Danielsson and Richard [6] and Sandmann and Koopman [43] suggest to approximate the likelihood function via importance sampling techniques. Finally Shephard [44] propose a simulated EM approach.
Let us consider the stochastic volatility model (35). 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 startup values of the algorithm are much better with the normal approximation. We found indeed that the starting point 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 (24) was much closer to the starting value $\bar{\theta}$ than to the true value of θ. Replications of the method with solutions of the previous step as new starting values, as suggested in Geyer [14] did not necessarily result in improved estimates, although the method usually converged to a fixed point. We therefore chose to initialize 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:

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

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

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

This property considerably improves both computing time and storage problems.
Figure 5 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 very important to point out that, as in the factor ARCH example, this estimation result have been obtained with a very small sample size ($T = 100$).

	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 (35) with a simulated sample of size $T = 100$ and a Bayes estimate as starting value.

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, since it applies to most latente 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

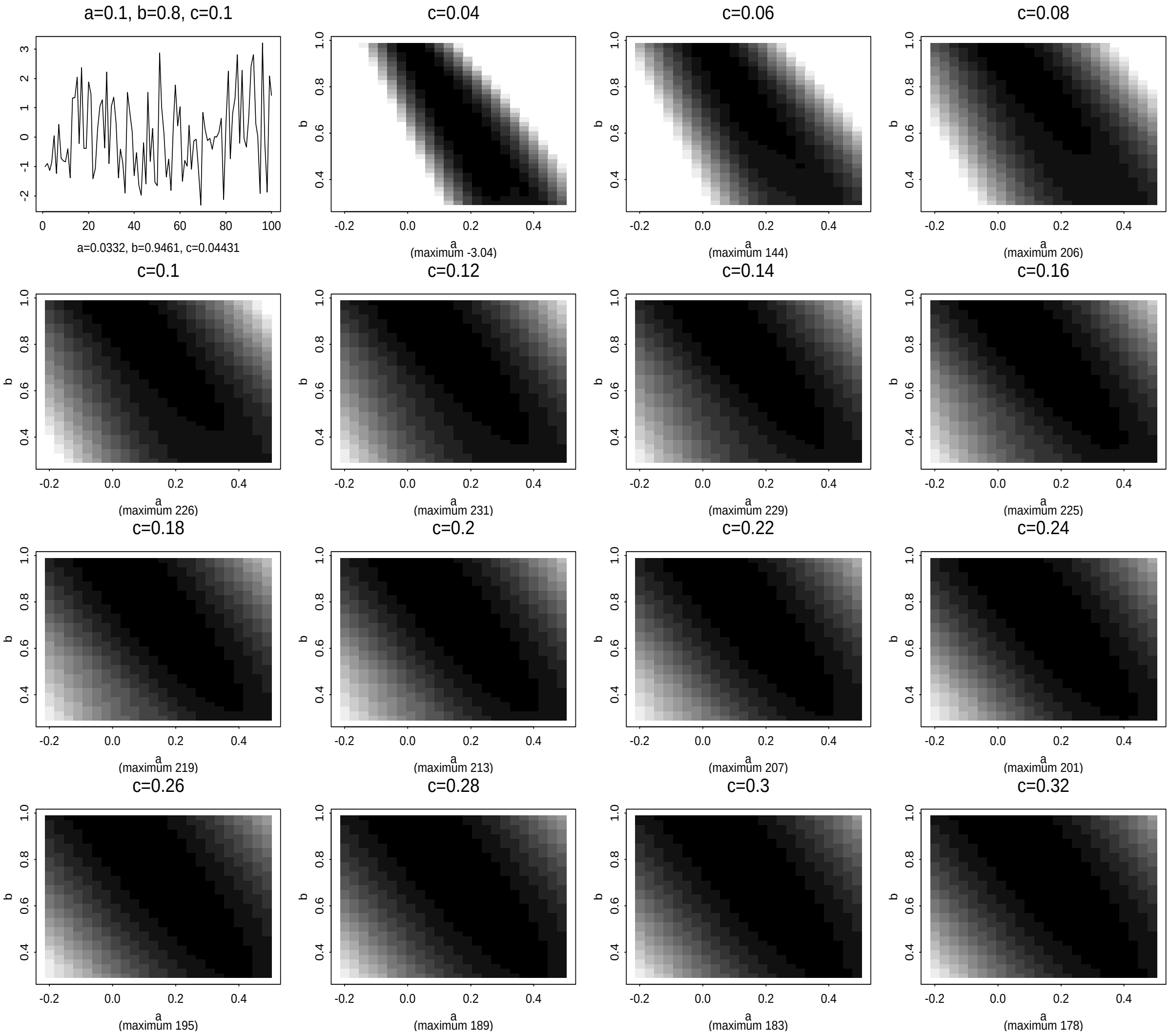

Figure 5: Evaluated slices in c of the log-likelihood ratio associated with (35) for a simulated sample of size $T = 100$ and 5000 iterations of the completion step. See Table 10 for the parameter values.

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 starting point may be important in some examples.

Appendix 1

Proof of Proposition 1

Lemma
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 the Lemma
$I(h)$ can be computed by successive integrations with respect to 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 with respect to $d\mu_1^{y_1^*}(y_1)$ is done, we have to integrate the function thus obtained with respect to the product measure $d\mu_{*1}(y_1^*) \otimes d\mu_2^{y_2^*}(y_2)$. Using the Fubini's theorem we know that the order of integration can be inverted and, integrating first with respect to $d\mu_2^{y_2^*}(y_2)$ we get a function which must be integrated with respect to the product measure $d\mu_{*1}(y_1^*) \otimes d\mu_{*2}(y_2^*) \otimes d\mu_3^{y_3^*}(y_3)$. Using again the Fubini's theorem we can integrate first with respect to $d\mu_3^{y_3^*}(y_3)$. Iterating this procedure we see that $I(h)$ can be written

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

□

Let us now consider a Borel set of $\mathbb{R}^{Tp^*}$, denoted by A. We have

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

where P^Z is the distribution of the random variable Z. Using the previous lemma we get:

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

This shows that $P_\theta^{y^{*T}}$ is absolutely continuous with respect to $\otimes_{t=1}^T \mu_{*t}(y_t^*)$, with a p.d.f. given by:

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

Similarly we have:

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

and formulae (8) are established.

Now let A and B be Borel sets of $I\!R^{p^*T}$ nd $I\!R^{pT}$ respectively. We have, using the lemma:

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

This shows that the conditional distribution of y^T given y^{*T} is absolutely continuous with respect to the measure $\otimes_{t=1}^T \mu_t^{y_t^*}(y_t)$ and that its p.d.f. is

$$\frac{f(y^{*T},y^T;\theta)}{f(y^{*T};\theta)}.$$

The symmetrical result is obvious and formulae (9) are established.

Appendix 2

Proof of Proposition 2

Note first that from the lemma we have for any Borel set $A_k \subset I\!R^{p^*(T-k)}$, $B \subset I\!R^{pT}$

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

This shows that $P_\theta^{y^{*T-k},y^T}(A_k \times B)$ is absolutely continuous with respect to $\otimes_{t=1}^{T-k}\nu_t$ $\otimes_{t=T-k+1}^T \mu_t$ 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

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

Moreover, the expectation of Proposition 2 can be written

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

Appendix 3

The framework considered in this paper allows for the computation of various conditional p.d.f. For instance the conditional distribution of y_t^* given (y^{*t-1}, y^{t+1}) has a p.d.f. with respect to $\mu_{*t}^{y_t}$, which is:

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

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

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

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

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

and, therefore, the conditional distribution of y_t^* given (y^{*t-1}, y^{t+1}) has a p.d.f. with respect to $\mu_{*t}^{y_t}$ which is:

$$\frac{f(y^{*t}, y^{t+1};\theta)}{f(y^{*t-1}, y^{t+1};\theta)} = f_t(y_t^*|y^{*t-1}, y^{t+1};\theta) \qquad \text{(say)}.$$

From the form of these joint p.d.f. we get:

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

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

References

[1] Andersen, T.G. and Sorensen, B.E., GMM estimation fo a stochastic volatility model: a Monte Carlo study, *Journal of Business & Economic Statistics*, 14, 328-352, 1996.

[2] Albert, J.H. and Chib, S., Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts, *Journal of Business & Economic Statistics*, 11, 1-6, 1993.

[3] Billio, M. and Monfort, A., Switching state space models, Doc. travail CREST no. 9557, Insee, Paris, 1995 (forthcoming in *Journal of Statistical Planning and Inference*).

[4] Billio, M., Monfort, A. and Robert C.P., Bayesian estimation of switching ARMA models, Doc. travail CREST no. 9650, Insee, Paris, 1996.

[5] Celeux, G. and Diebolt, J., The SEM algorithm: a probabilistic teacher derived from the EM algorithm for the mixture problem. *Comp. Statist. Quater.*, 2, 73-82, 1985.

[6] Danielsson, J. and Richard, J.F., Accelerated Gaussian Importance Sampler with Application do Dynamic Latent Variable Models, *Journal of Applied Econometrics*, 8, 153-173, 1993.

[7] Diebold, D. and Nerlove, M., The dynamic of exchange rate volatility: a multivariate latent factor ARCH model, *Journal of Applied Econometrics*, 4, 1-22, 1989.

[8] Diebolt, J. and Ip, E.H.S., Stochastic EM: method and applications, in *Markov Chain Monte Carlo in Practice*, eds Gilks W.R. and Spiegelhalter D.J., Chapman and Hall, London, 1996.

[9] Dueker, M.J., Markov Switching in GARCH Processes and Mean-Reverting Stock Market Volatility, *Journal of Business & Economic Statistics*, 15/1, 26-34, 1997.

[10] Duffie, J. and Singleton, K., Simulated Moments Estimation of Markov Models of Asset Prices, *Econometrica*, 61, 929-952, 1993.

[11] Duflo, M., Random Iterative Models, in *Applications of Mathematics*, eds. Karatzas I. and Yor M., Vol. 34, Springer Verlag, 1997.

[12] Engle, R., Ng, V. and Rothschild, M., Asset pricing with a factor ARCH variance structure: empirical estimates for Treasury Bills, *Journal of Econometrics*, 45, 213-237, 1990.

[13] Geyer, C.J., On the convergence of Monte Carlo maximum likelihood caculations, *Journal of Royal Statistical Society*, (Ser. B) 56, 261-274, 1994.

[14] Geyer, C.J., Estimation and optimization of functions, in *Markov Chain Monte Carlo in Practice*, eds Gilks W.R. and Spiegelhalter D.J., Chapman and Hall, London, 1996.

[15] Geyer, C.J. and Thomson, E.A., Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *Journal of the Royal Statistical Society*, (Ser. B) 54, 657-699, 1992.

[16] Geweke, J., Antithetic acceleration of Monte Carlo integration in Bayesian inference, *Journal of Econometrics*, 38, 73-90, 1988.

[17] Geweke, J., Bayesian inference in econometric models using Monte Carlo integration, *Econometrica*, 57, 1317-1340, 1989.

[18] Geweke, J., Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints, *Computer Sciences and Statistics: Proc. 23d Symp. Interface*, 571-577, 1991 Smith A.F.M., 169-193, Oxford University Press, Oxford, 1992.

[19] Ghysels, E., Harvey A.C. and Renault, E. , 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, 1996.

[20] Gouriéroux, C. and Monfort, A., *Simulation Based Methods in Econometrics*, Oxford University Press, Oxford, 1996.

[21] Gouriéroux, C., Monfort, A. and Renault, E., Indirect Inference, *Journal of Applied Econometrics*, 8, 85-118, 1993.

[22] HAJIVASSILIOU, V.A. AND MCFADDEN, D., The method of simulated scores, with applications to models of external debt crises, *Econometrica*, 64, 1996.

[23] HAMILTON, J.D., Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates, *Journal of Economic Dynamics and Control*, 12, 385-423, 1988.

[24] HAMILTON, J.D., A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57 (2), 357-384, 1989.

[25] HARVEY, A.C., RUIZ, E. AND SHEPHARD, N., Multivariate stochastic variance modles, *Review of Economic Studies* 61, 247-264, 1994.

[26] HEITJAN, D.F. AND RUBIN, D.B., Ignorability and coarse data, *Annals of Statistics*, 19, 2244-2253, 1991.

[27] JACQUIER, E., POLSON, N.G. AND ROSSI, P.E., Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics*, 12, 371-417, 1994.

[28] KIM, C.J., Dynamic Linear Models With Markov Switching, *Journal of Econometrics* , 64, 1-22, 1994.

[29] KIM, S., SHEPHARD, N. AND CHIB, S., Stochastic volatility: likelihood inference and comparison with ARCH models, *mimeo*, forthcoming *Review of Economic Studies*, 1998.

[30] LAM, P.S., The Hamilton model with a general autoregressive component, *Journal of Monetary Economics*, 26, 409-432, 1990.

[31] LAROQUE, G. AND SALANIÉ, B., Simulation based estimation of models with lagged latent variables, *Journal of Applied Econometrics*, 8, 119-133, 1993.

[32] LEE, L.F., Simulation estimation of dynamic switching regression and dynamic disequilibrium models - some Monte Carlo results, *Journal of Econometrics*, 78, 179-204, 1997.

[33] LEE, L.F., A smooth likelihood simulator for dynamic disequilibrium models, *Journal of Econometrics*, 78, 257-294, 1997.

[34] LITTLE, R.J.A. AND RUBIN, D.B., On jointly estimating parameters and missing data by maximizing the complete-data likelihood, *American Statist.*, 37/3, 218-220, 1983.

[35] LIU, J.S., WONG, W.H. AND KONG, A., Correlation structure and convergence rate of the Gibbs sampler with various scans, *Journal of Royal Statistical Society*, (Ser. B) 57, 157-169, 1995.

[36] MADDALA, G.S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.

[37] MANRIQUE, A. AND SHEPHARD, N., Simulation based likelihood inference for limited dependent processes, mimeo 1996.

[38] McFadden, D., A method of simulated moments for estimation of discrete response models without numerical integration, *Econometrica*, 57, 995-1026, 1989.

[39] Monfardini, C., Estimating stochastic volatility models through indirect inference, *Working Paper n.* 96/22, European University Institute, 1996.

[40] Philippe, A., Simulation of right- and left-truncated Gamma distributions by mixtures, *Statistics and Computing*, 7, 173-182, 1997.

[41] Robert, C.P., *Méthodes de Monte Carlo par Chaînes de Markov*, Economica, Paris, 1996.

[42] Robert, C.P. and Titterington, M., Resampling schemes for hidden Markov models and their application for maximum likelihood estimation, *Statistics and Computing* (to appear), 1997.

[43] Sandmann, G. and Koopman, S.J., Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *mimeo*, London School of Economics, 1997.

[44] Shephard, N., Fitting non-linear time series models, with applications to stochastic variance models, *Journal of Applied Econometrics*, 8, 135-152, 1993.

[45] Shephard, N., Partial non-Gaussian state space, *Biometrika*, 81, 115-131, 1994.

[46] Shephard, N., 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, 1996.

[47] Shephard, N. and Pitt, M.K., Likelihood analysis of non-Gaussian measurement time series, *Biometrika*, 84, 653-668, 1997.

[48] Taylor, S.J., Modeling stochastic volatility: a review and comparative study, *Mathematical Finance*, 4, 183-204, 1994.

[49] Tierney, L., Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, 22, 1701-1786, 1994.

[50] Wei, G.C.G. and Tanner, M.A., A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm, *Journal of the American Statistical Association*, 85, 699-704, 1990.