ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Efficient Gibbs sampling for Markov switching GARCH models☆

Monica Billio [a], Roberto Casarin [a], Anthony Osuntuyi [a,b,*]

[a] Department of Economics, University Ca' Foscari of Venice, Fondamenta San Giobbe 873, 30121, Venice, Italy
[b] Department of Mathematics, Obafemi Awolowo University, 220005, Ile-Ife, Nigeria

ARTICLE INFO

Article history:
Received 4 September 2013
Received in revised form 14 April 2014
Accepted 14 April 2014
Available online xxxx

Keywords:
Bayesian inference
GARCH
Markov-switching
Multiple-try Metropolis

ABSTRACT

Efficient simulation techniques for Bayesian inference on Markov-switching (MS) GARCH models are developed. Different multi-move sampling techniques for Markov switching state space models are discussed with particular attention to MS-GARCH models. The multi-move sampling strategy is based on the Forward Filtering Backward Sampling (FFBS) approach applied to auxiliary MS-GARCH models. A unified framework for MS-GARCH approximation is developed and this not only encompasses the considered specifications, but provides an avenue to generate new variants of MS-GARCH auxiliary models. The use of multi-point samplers, such as the multiple-try Metropolis and the multiple-trial metropolized independent sampler, in combination with FFBS, is considered in order to reduce the correlation between successive iterates and to avoid getting trapped by local modes of the target distribution. Antithetic sampling within the FFBS is also suggested to further improve the sampler's efficiency. The simulation study indicates that the multi-point and multi-move strategies can be more efficient than other MCMC schemes, especially when the MS-GARCH is not strongly persistent. Finally, an empirical application to financial data shows the efficiency and effectiveness of the proposed estimation procedure.

1. Introduction

The modeling of financial market volatility remains a prominent area of research in finance given the important role it plays in a variety of financial problems such as asset pricing and risk management. Among volatility models, the Bollerslev (1986) Generalized Autoregressive Conditional Heteroskedastic (GARCH) model and its variants rank as the most popular class of models among researchers and practitioners (e.g., see Ardia (2008)). However, empirical studies document that this class of models exhibits high persistence in its conditional variance, i.e. the variance process is close to being non-stationary (nearly integrated). Lamoureux and Lastrapes (1990), among others, argue that the presence of structural changes in the variance process, for which standard GARCH processes cannot account for, may be responsible for this phenomenon. To buttress this point, Mikosch and Starica (2004) estimate a GARCH model on a sample that exhibits structural changes in its conditional variance and obtained a nearly integrated GARCH effect from the estimates. Based on this observation, Hamilton and Susmel (1994) and Cai (1994) propose a Markov Switching-Autoregressive Conditional Heteroskedastic (MS-ARCH)

☆ The Supplementary Material presents the derivation of the proposal distributions for the MS-GARCH parameters and graphical illustration of the results of the algorithm comparison (see Appendix A).

* Corresponding author at: Department of Economics, University Ca' Foscari of Venice, Fondamenta San Giobbe 873, 30121, Venice, Italy.
E-mail addresses: billio@unive.it (M. Billio), r.casarin@unive.it (R. Casarin), ayokunlet@yahoo.com (A. Osuntuyi).

http://dx.doi.org/10.1016/j.csda.2014.04.011
0167-9473/

model, governed by a state variable that follows a first-order Markov chain and captures the high volatility persistence. Gray (1996) instead considers a Markov switching GARCH (MS-GARCH) model since it can be written as an infinite order ARCH model and may be more parsimonious than the MS-ARCH model.

The MS-GARCH class of models is gradually becoming a workhorse among economist and financial practitioners for analyzing financial market data (e.g., see Marcucci (2005)). For the practical implementation of this class of models, it is crucial to have reliable parameter estimators. The Maximum Likelihood (ML) approach is a natural route to parameter estimation. However, the ML technique is not computationally feasible for MS-GARCH models because of the path dependence problem (see Gray (1996)). To this end, Henneke et al. (2011) and Bauwens et al. (2010) propose a Bayesian approach based on Markov Chain Monte Carlo (MCMC) for estimating parameters of Markov Switching-Autoregressive Moving Average-Generalized Autoregressive Conditional Heteroskedastic (MS-ARMA–GARCH) and MS-GARCH models, respectively. Their proposed algorithm samples each state variable individually given the others (single-move Gibbs sampler). This sampler shows slow convergence and it is computationally demanding. Thus, great attention has been paid to improving such inefficiencies in the context of continuous and possibly non-Gaussian and nonlinear state space models. See, for example, Frühwirth-Schnatter (1994), Koopman and Durbin (2000), De Jong and Shephard (1995) and Carter and Kohn (1994) for multi-move Gibbs sampling, and So (2006) for multi-points and multi-move Gibbs sampling schemes for continuous and nonlinear state space models. To the best of our knowledge, there are few papers on efficient multi-move sampling schemes for discrete or mixed state space models in existence. See Kim and Nelson (1999) for a review of multi-move Gibbs for conditionally linear models, Billio et al. (1999) for global Metropolis–Hastings algorithm for sampling the hidden states of MS-ARMA models and Fiorentini et al. (2012) for multi-move sampling in dynamic mixture models. In regard to MS-GARCH models, Ardia (2008) develops a Gibbs sampler for the state variables of the model of Haas et al. (2004), which is a different representation of the MS-GARCH model; He and Maheu (2010) propose a Sequential Monte Carlo (SMC) algorithm for GARCH models subject to structural breaks; Bauwens et al. (2014) propose a Particle MCMC (PMCMC) algorithm for estimating GARCH models subject to either structural breaks or regime switching. Elliott et al. (2012) propose instead a Viterbi-based technique, for sampling the hidden state variables of a MS-GARCH model. Finally, Dufays (2012) proposes an infinite MS-GARCH model and develops, in a parallel and independent research, a Metropolis–Hastings algorithm with proposal distribution based on Klaassen's approximation. However, a general approach to the design of efficient proposal distributions for sampling the hidden states of a MS-GARCH, which is one of the main contributions of this paper, is not discussed in Dufays (2012). Alternative estimation procedures such as a generalized method of moments (GMM) (Francq and Zakoian, 2008) and neural networks (Bildirici and Ersin, 2012) have also been proposed in the literature.

In this paper, we develop an efficient simulation-based estimation approach for MS-GARCH models with a finite number of regimes wherein the conditional mean and conditional variance of the observation process may change over time. We follow a data augmentation framework by incorporating the state variables into the parameter vector. In particular, we propose a Bayesian approach based on a MCMC algorithm which allows us to circumvent the problem of path dependence by simultaneously generating the states (via the multi-move Gibbs sampler) from their joint distribution. Our strategy for sampling the state variables is based on Forward Filtering Backward Sampling (FFBS) techniques. As in the case of mixed hidden state models, the FFBS algorithm cannot be applied directly to switching GARCH models. Accordingly, we develop a unified framework of MS-GARCH approximation based on a regime-collapsing procedure. This framework not only encompasses specifications already existing in the literature, but also enables us to generate new auxiliary models. Using this framework, we propose and discuss different multi-move sampling techniques for a MS-GARCH model by applying FFBS to an auxiliary MS-GARCH model.

Another original contribution of the paper relates to the Metropolis step for the hidden states. To efficiently estimate MS-GARCH models, we consider the class of generalized (multi-point) Metropolis algorithms (see Liu (2002), Chapter 5) that extends the standard Metropolis–Hastings (MH) approach (Hastings, 1970) and (Metropolis et al., 1953). See Liu (2002) and Robert and Casella (2010) for an introduction to MH algorithms and a review of various extensions. Multi-point samplers have been proved, both theoretically and computationally, to be effective in improving the mixing rate of the MH chain and the efficiency of the Monte Carlo estimates based on the output of the chain. The main feature of the multi-point samplers is that at each iteration of the MCMC chain the new value of the chain is selected among multiple proposals, while in the MH algorithm one accepts or rejects a single proposal. Another strength of the multi-point scheme is that parallelization can easily be introduced, along the lines described in Casarin et al. (2013), thus possibly leading to a better efficiency of the sampling. In this paper, we apply the multiple-try Metropolis (MTM) (see (Liu et al., 2000)) and some modified MTM algorithms. The superiority of the MTM over standard MH has been proven in Craiu and Lemieux (2007), who also propose to apply antithetic and quasi-Monte Carlo techniques to obtain good proposals in the MTM. So (2006) applies MTM to the estimation of latent-variable models and finds evidence of better efficiency of the MTM over standard MH samplers for the latent variable estimation. The author also finds that the efficiency of MTM can be further increased by the use of multi-move sampling. Casarin et al. (2013) apply the MTM transition within the context of interacting chains and they provide a comparison with standard interacting MH. They also estimate the efficiency gain when using interacting MTM combined with block-sampling for the estimation of stochastic volatility models. Thus, we combine the MTM sampling strategies with the approximated FFBS techniques for the Markov switching process. In this sense, we extend the work of So (2006) to the more complex case of Markov-switching nonlinear state space models. In fact, the use of multiple proposals is particularly suited within this context, where the forward filter is used at each iteration to generate only one proposal with a large computational cost. The use of multiple proposals based on the same run of the forward filter is discussed. We

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

also apply to this context the antithetic sampling technique proposed by Craiu and Lemieux (2007) to generate negatively correlated proposals within the multiple-try algorithm, and suggest a Forward Filtering Backward Antithetic Sampling (FFBAS) algorithm which combines the permuted displacement algorithm of Craiu and Meng (2005) with the FFBS and possibly produces pairwise negative association among the trajectories of the hidden states. Note that our approach could easily be extended to other discrete or mixed-state space models. Finally, our efficient sampling procedure may also be applied to simulation-based maximum likelihood context (such as Billio et al. (1998) and Augustyniak (2013)) to make inferences on the MS-GARCH model.

The paper is organized as follows. Section 2 introduces the MS-GARCH model and discusses the inference difficulties with the estimation methods proposed in the literature. In Section 3, we present a Bayesian inference approach and introduce multi-move multi-point sampling strategies. In Section 4, we study the efficiency of our estimation procedure through some simulation experiments. In Section 5 we provide an application to S&P 500 return series while Section 6 concludes and discusses possible extensions.

2. Markov switching GARCH models

2.1. The model

A Markov switching GARCH model is a nonlinear specification of the evolution of a time series which is assessed to be affected by different states of the world and for which the conditional variance in each state follows a GARCH process. More specifically, let y_t be the observed variable (e.g. the return on some financial asset) and s_t a discrete, unobserved, state variable which could be interpreted as the state of the world at time t. Define $(y_t, \ldots, y_u)$ and $(s_t, \ldots, s_u)$ as $y_{t:u}$ and $s_{t:u}$ respectively, whenever $t \leq u$. Then an MS-GARCH model can be defined as follows

$$y_t = \mu_t(y_{1:t-1}, \theta_\mu(s_t)) + \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{1}$$

$$\sigma_t^2(y_{1:t-1}, \theta_\sigma(s_t)) = \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2(y_{1:t-2}, \theta_\sigma(s_{t-1})), \tag{2}$$

where, $\mathcal{N}(0,1)$ denotes the standard normal distribution, θ_μ represents the parameters of the conditional mean of the returns series, $\epsilon_t = \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t, \theta_\sigma(s_t) = (\gamma(s_t), \alpha(s_t), \beta(s_t)), \gamma(s_t) > 0, \alpha(s_t) \geq 0, \beta(s_t) \geq 0$, and $s_t \in \{1, \ldots, M\}$, $t = 1, \ldots, T$, is assumed to follow a M-state first order Markov chain with transition probabilities $\{\pi_{ij}\}_{i,j=1,2,\ldots,M}$:

$$\pi_{ij} = p(s_t = i | s_{t-1} = j, \theta_\pi), \qquad \sum_{i=1}^{M} \pi_{ij} = 1 \quad \forall j = 1, 2, \ldots, M,$$

where θ_π represents the parameters of the transition matrix. The parameter shift functions $\gamma(s_t)$, $\alpha(s_t)$ and $\beta(s_t)$, describe the dependence of parameters on the regime, s_t i.e.

$$\gamma(s_t) = \sum_{m=1}^{M} \gamma_m \mathbb{I}_{\{s_t=m\}}, \qquad \alpha(s_t) = \sum_{m=1}^{M} \alpha_m \mathbb{I}_{\{s_t=m\}}, \quad \text{and} \quad \beta(s_t) = \sum_{m=1}^{M} \beta_m \mathbb{I}_{\{s_t=m\}},$$

where,

$$\mathbb{I}_{\{s_t=m\}} = \begin{cases} 1, & \text{if } s_t = m, \\ 0, & \text{otherwise.} \end{cases}$$

By defining the allocation variable, s_t, as an M-dimensional discrete vector, $\xi_t = (\xi_{1t}, \ldots, \xi_{Mt})'$, where $\xi_{mt} = \mathbb{I}_{\{s_t=m\}}$, $m = 1, \ldots, M$, Eqs. (1) and (2) can be written compactly as

$$y_t = \mu_t(y_{1:t-1}, \xi_t'\theta_\mu) + \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{3}$$

$$\sigma_t^2(y_{1:t-1}, \xi_t'\theta_\sigma) = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon_{t-1}^2 + (\xi_t'\beta)\sigma_{t-1}^2(y_{1:t-2}, \xi_{t-1}'\theta_\sigma), \tag{4}$$

where $\epsilon_t = \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t$, $\gamma = (\gamma_1, \ldots, \gamma_M)'$, $\alpha = (\alpha_1, \ldots, \alpha_M)'$, $\beta = (\beta_1, \ldots, \beta_M)'$, $\theta_\mu = (\theta_{1\mu}, \ldots, \theta_{M\mu})'$ and $\theta_\sigma = (\theta_{1\sigma}, \ldots, \theta_{M\sigma})'$ with $\theta_{m\sigma} = (\gamma_m, \alpha_m, \beta_m)'$ for $m = 1, \ldots, M$. for $t = 1, \ldots, T$. Let $\pi = (\pi_1, \ldots, \pi_M)$, with $\pi_i = (\pi_{i1}, \ldots, \pi_{iM})$ for $i = 1, 2, \ldots, M$ and $\sum_{i=1}^{M} \pi_{ij} = 1$ for all $j = 1, 2, \ldots, M$. Since ξ_t follows an M-state first order Markov chain, we define the transition probabilities $\{\pi_{ij}\}_{i,j=1,2,\ldots,M}$ by

$$\pi_{ij} = p(\xi_t = e_i | \xi_{t-1} = e_j, \theta_\pi),$$

where e_i is the ith column of an M-by-M identity matrix. The conditional probability of ξ_t given ξ_{t-1}, and θ_π is given by

$$p(\xi_t | \xi_{t-1}, \theta_\pi) = \prod_{m=1}^{M} (\pi_m \xi_{t-1})^{\xi_{mt}}, \tag{5}$$

which implies that the probability with which event m occurs at time t is $\pi_m \xi_{t-1}$.

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

2.2. Inference issues

Estimating Markov switching GARCH models is a challenging problem since the likelihood of y_t depends on the entire sequence of past states up to time t due to the recursive structure of the volatility. To elaborate on this, the likelihood function of the switching GARCH model is given by

$$\mathcal{L}(\theta|y_{1:T}) \equiv f(y_{1:T}|\theta) = \sum_{i_1=1}^{M} \dots \sum_{i_T=1}^{M} f(y_{1:T}, \xi_1 = e_{i_1}, \dots, \xi_T = e_{i_T}|\theta), \tag{6}$$

where $\theta = (\{\theta_{m\mu}, \theta_{m\sigma}\}_{m=1,\dots,M}, \theta_\pi)$. Setting $\xi_{s:t} = (\xi_s', \dots, \xi_t')$ whenever $s \le t$, the joint density function of $y_{1:t}$ and $\xi_{1:t}$ on the right hand side of Eq. (6) is

$$\begin{aligned} f(y_{1:T}, \xi_{1:T}|\theta) &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) p(\xi_t|\xi_{1:t-1}, \theta_\pi) \\ &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \left(\prod_{i=1}^{M} (\pi_i \xi_{t-1})^{\xi_{it}} \right), \end{aligned} \tag{7}$$

with,

$$f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \propto \frac{1}{\sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)} \exp\left(-\frac{1}{2} \left(\frac{y_t - \mu_t(y_{1:t-1}, \xi_t'\theta_\mu)}{\sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)} \right)^2 \right).$$

Given σ_1, recursive substitution in Eq. (4) yields

$$\sigma_t^2 = \sum_{i=0}^{t-2} \left[\xi_{t-i}'\gamma + (\xi_{t-i}'\alpha)\epsilon_{t-1-i}^2 \right] \prod_{j=0}^{i-1} \xi_{t-j}'\beta + \sigma_1^2 \prod_{i=0}^{t-2} \xi_{t-i}'\beta. \tag{8}$$

Eq. (8) shows the dependence of the conditional variance at time t on the entire history of the switching process (path dependence problem). The evaluation of the likelihood function over a sample of length T (see Eq. (6)) involves the integration (summation) over all M^T unobserved states (i.e. integration over all M^T possible (unobserved) regime paths). This requirement makes the maximum likelihood estimation of θ infeasible in practice.

Two major approaches have been developed in the literature in order to circumvent this path dependence problem. One approach involves the use of model approximations, while the other is simulation based.

In regard to the model approximation approach, Cai (1994) and Hamilton and Susmel (1994) assumed an MS-ARCH model. This approximation makes the model tractable because it removes the lagged conditional variance from the conditional variance equation, thus avoiding the path dependence problem. Kaufman and Frühwirth-Schnatter (2002) employed the algorithm developed by Chib (1996) for Markov mixture models to compute the marginal likelihood of the MS-ARCH model, but noted that this methodology cannot be carried over to the MS-GARCH model because of the path dependence problem. Another approximation approach can be credited to Gray (1996) who noted that the conditional density of the return is a mixture of distributions with time-varying mixing parameters. Moreover, under the normality assumption, Gray (1996) suggests to approximate the conditional variance by averaging the regime-specific conditional variances. Extensions of Gray's model can be found in Dueker (1997), Klaassen (2002) and Haas et al. (2004) among others. Abramson and Cohen (2007) provide stationarity conditions for some of these approximations. The problem with this approach is that these approximations cannot be verified.

Among the simulation based approaches proposed in the literature, there is the Bayesian estimation technique proposed by Bauwens et al. (2010). In particular, they develop a single-move MCMC Gibbs sampler for a Markov switching GARCH model with a fixed number of regimes. The authors also provide sufficient conditions for geometric ergodicity and the existence of moments of the process. Their estimation approach, though quite promising, has one main limitation that has rendered it unattractive: the single-move Gibbs sampler is inefficient (i.e. draws from the single-move scheme are known to be highly correlated and thus slow down the convergence of the Markov chain). Alternative simulation based approaches rely on particle filters also known as sequential Monte Carlo methods. He and Maheu (2010) develop a sequential Monte Carlo method for estimating GARCH models subject to an unknown number of structural breaks. Bauwens et al. (2014) propose a particle MCMC approach to GARCH models subject to structural breaks or regime switching. In a parallel research study to ours, but independent of it, Dufays (2012) develops an MCMC scheme based on Klaassen's approximation for block sampling of the hidden state of an infinite state MS-GARCH model. Besides, we introduce a general approach based on conditional expectation to generate auxiliary models; thus Klaassen's approximation is simply a specific case of our general approach. Augustyniak (2013) develops hybrid Monte Carlo expectation–maximization and Monte Carlo maximum likelihood algorithms to calculate the maximum likelihood estimator for the MS-GARCH model.

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

3. Bayesian inference

Based on the aforementioned inference issues associated with MS-GARCH models, we present a Bayesian approach based on an MCMC Gibbs algorithm which allows us to circumvent the path dependence problem and efficiently sample the hidden state trajectory. The algorithm generates samples from the posterior distribution which are then used for its characterization. We follow a data augmentation framework by treating the state variables as parameters of the model and construct the likelihood function assuming that the states are known.

Before proceeding with the elicitation of our proposed Bayesian technique, it is important that we make explicit the parametric specification of the conditional mean, $\mu_t(y_{1:t-1}, \xi_t'\theta_\mu)$, of the return process y_t in Eq. (3). Since our major aim is to define a technique for sampling the state variables efficiently, which in turn will affect other parameter estimates, we assume for expository purpose a conditional mean defined by a switching parameter (i.e. $\xi_t'\mu$ where $\mu = (\mu_1, \ldots, \mu_M)'$). Alternative specifications such as the switching ARMA process could be thought of for the conditional mean and time varying transition probabilities may be defined in place of constant transition probabilities by following Gray (1996) approach (i.e. specifying the transition probabilities as a function of past observables). Under this specification, the augmented parameters set by our model consists of $\xi_{1:T}$ and $\theta = (\theta_\mu, \theta_\sigma, \theta_\pi)$, where $\theta_\mu = \mu$, $\theta_\pi = (\{\pi_m\}_{m=1,\ldots,M})$ and $\theta_\sigma = (\{\theta_{m\sigma}\}_{m=1,\ldots,M})$, with $\theta_{m\sigma} = (\gamma_m, \alpha_m, \beta_m)$, $\pi_m = (\pi_{m1}, \ldots, \pi_{mM})$ and $\sum_{m=1}^{M} \pi_{mm^*} = 1, \forall\, m^* = 1, \ldots, M$. We assume a fairly informative prior distribution for the transition probabilities θ_π

$$\theta_\pi \sim \prod_{m=1}^{M} \text{Dirichlet}(\nu_{1m}, \ldots, \nu_{Mm}),$$

where $\nu_{1m}, \ldots, \nu_{Mm}, m = 1, \ldots, M$ are known hyperparameters. Other parameters of the MS-GARCH model are assumed to follow independent uniform priors. If the uniform distribution is defined on a bounded domain (see Bauwens et al. (2010)) then the posterior is always defined. If the uniform distribution is an improper prior, the posteriors are not always proper distributions (see Wasserman (2000)). The posterior of the kth regime parameter is not proper if there are no observations allocated to the kth regime. It is possible to avoid this offensive grouping of the data by rejecting, at each iteration of the Gibbs sampler, the draws of the sequence of allocation variables, ξ_t, $t = 1, \ldots, T$, that do not belong to the set $\mathcal{S} = \{\xi_{1:T} : \sum_{t=1}^{T} \xi_{jt} \geq 1, \forall j = 1, \ldots, M\}$. Moreover, to avoid the label switching problem, we assume the identifiability constraint: $\gamma_1 < \cdots < \gamma_M$. See Frühwirth-Schnatter (2006) for an introduction to the label switching problem for dynamic mixtures and MS models and Bauwens et al. (2010) for an illustration of the identification constraint for MS-GARCH models. As in the case of mixture models, the label switching problem is indeed a major problem. Thus, a wrong identification constraint can bias the results of our estimation exercise. The joint prior distribution is thus proportional to

$$f(\theta) \propto \prod_{m=1}^{M} \text{Dirichlet}(\nu_{1m}, \ldots, \nu_{Mm}) \mathbb{I}_{\{\gamma_1 < \cdots < \gamma_M\}}. \tag{9}$$

The posterior density of the augmented parameter vector given by

$$f(\theta, \xi_{1:T}|y_{1:T}) \propto f(y_{1:T}|\xi_{1:T}, \theta) f(\xi_{1:T}|\theta) f(\theta), \tag{10}$$

cannot be identified with any standard distribution; hence, we cannot sample directly from it. Using the Gibbs sampler, we can generate samples from this high-dimensional posterior density. This will be done by iteratively sampling from the following three full conditional distributions

1. $p(\xi_{1:T}|\theta, y_{1:T})$,
2. $f(\theta_\pi|\theta_\mu, \theta_\sigma, \xi_{1:T}, y_{1:T}) = f(\theta_\pi|\xi_{1:T})$,
3. $f(\theta_\sigma, \theta_\mu|\theta_\pi, \xi_{1:T}, y_{1:T}) = f(\theta_\sigma, \theta_\mu|\xi_{1:T}, y_{1:T})$.

These distributions are easy to manage and sample from because they can either be associated with a known distribution or simulated by a lower dimensional auxiliary sampler. In the following subsections, we present in detail our sampling procedure for both the single-move and multi-move algorithms.

3.1. Sampling the state variables $\xi_{1:T}$

To sample $\xi_{1:T}$ using the single-move algorithm, one relies on computing

$$p(\xi_t|\xi_{1:t-1}, \xi_{t+1:T}, \theta, y_{1:T}) \propto \prod_{m=1}^{M} (\pi_m \xi_{t-1})^{\xi_{mt}} (\pi_m \xi_t)^{\xi_{m,t+1}} \prod_{j=t}^{T} f(y_j|\xi_j, \theta, y_{1:j-1}), \tag{11}$$

for each value ξ_t in $\{e_m : m = 1, \ldots, M\}$. Sampling from such a distribution, once the probabilities are known, is similar to sampling from a Multinomial distribution. On the other hand, the full joint conditional distribution of the state variables, $\xi_{1:T}$, given the parameter values and return series

$$p(\xi_{1:T}|\theta, y_{1:T}) \propto f(y_{1:T}|\xi_{1:T}, \theta) p(\xi_{1:T}|\theta), \tag{12}$$

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

is a non-standard. For this reason, we propose a generalized Metropolis (i.e. multi-point Metropolis) algorithm for generating the state variables. Multi-point samplers are designed to consider multiple proposals at each iteration of a Metropolis and to choose the new value of the chain from this trial set. The multi-move and multi-point sampling procedures are of interest because of their potential for addressing issues associated with the multi-modality of the target function and the autocorrelation of samples from the Metropolis chain. Our scheme generally involves running a FFBS on the auxiliary model to generate several proposals at each iteration step. Let the proposal distribution be denoted by

$$q(\xi_{1:T}|\theta, y_{1:T}) = q(\xi_T|\theta, y_{1:T}) \prod_{t=1}^{T-1} q(\xi_t|\xi_{t+1}, \theta, y_{1:t}), \tag{13}$$

where $q(\xi_t|\xi_{t+1}, \theta, y_{1:t}) \propto q(\xi_t|y_{1:t}, \theta) q(\xi_{t+1}|\xi_t, \theta)$ with $q(\xi_t|y_{1:t}, \theta)$ representing the filtered probability. A discussion of the proposal distribution is presented in Section 3.2. In the following, we discuss the three multi-point algorithms considered in this paper.

Algorithm 1 Multiple Try Metropolis Sampler (MTM)

i. Let $\xi_{1:T}^{(r-1)}$ be the value of the MTM at the $(r-1)$-th iteration.
ii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ containing K independent paths of the state variable drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$.
iii. Evaluate

$$W_k(\xi_{1:T,k}) = \frac{p(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T})}{q(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T},)}, \quad \forall\, k = 1, \ldots, K.$$

iv. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,k})}{\sum_{k=1}^{K} W_k(\xi_{1:T,k})}, \quad \forall\, k = 1, \ldots, K.$$

v. Construct a reference set $\{\xi_{1:T,1}^*, \xi_{1:T,2}^*, \ldots, \xi_{1:T,K}^*\}$ by setting the first $K-1$ elements to a new set of samples drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$ and the K-th element $\xi_{1:T,K}^*$ to $\xi_{1:T}^{(r-1)}$.
vi. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi_{1:T}^{(r)} = \begin{cases} \tilde{\xi}_{1:T}, & \text{if } u \le \alpha_{MTM}, \\ \xi_{1:T}^{(r-1)}, & \text{otherwise,} \end{cases}$$

where,

$$\alpha_{MTM} = \min\left(1, \frac{\sum_{k=1}^{K} W_k(\xi_{1:T,k})}{\sum_{k=1}^{K} W_k(\xi_{1:T,k}^*)}\right).$$

3.1.1. Multiple-try Metropolis sampler

Liu et al. (2000) suggest the multiple-try Metropolis (MTM) which considers several proposals so that a larger region of the sample space can possibly be visited. By using the multiple-try strategy, it is easier for the MTM chain to jump from one local mode to another and thus to converge quickly to the desired target distribution. In our MTM algorithm, samples from the proposal distribution will be generated by a FFBS algorithm. Below, we present a sketch of the main ingredients needed in both the Forward Filter (FF) and Backward Sampling (BS) steps of the algorithm and refer the reader to Frühwirth-Schnatter (2006) for a detailed presentation of this procedure. At time t, given θ and $y_{1:t}$ the FF probabilities are obtained by first computing the one-step ahead prediction

$$q(\xi_t|\theta, y_{1:t-1}) = \sum_{i=1}^{M} \left(\prod_{j=1}^{M} (\pi_j e_i)^{\xi_{j,t}} \right) q(\xi_{t-1} = e_i|\theta, y_{1:t-1}),$$

and then, the FF is

$$q(\xi_t|\theta, y_{1:t}) = \frac{g(y_t|\xi_t, \theta, y_{1:t-1}) q(\xi_t|\theta, y_{1:t-1})}{\sum_{i=1}^{M} g(y_t|\xi_t = e_i, \theta, y_{1:t-1}) q(\xi_t = e_i|\theta, y_{1:t-1})}, \tag{14}$$

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

where $g(y_t|\xi_t, \theta, y_{1:t-1})$ is the conditional density of the return process under the auxiliary model. Using the output of the FF, we compute $q(\xi_T|\theta, y_{1:T})$ and

$$q(\xi_t|\xi_{t+1}, \theta, y_{1:t}) = \frac{\prod_{j=1}^{M} \left(\pi_j \xi_t\right)^{\xi_{j,t+1}} q(\xi_t|\theta, y_{1:t})}{q(\xi_{t+1}|\theta, y_{1:t})}, \tag{15}$$

for $t = T-1, T-2, \ldots, 2, 1$. Then, at each time step we sample ξ_T from $q(\xi_T|\theta, y_{1:T})$ and ξ_t from $q(\xi_t|\xi_{t+1}, \theta, y_{1:t})$ iteratively for $t = T-1, T-2, \ldots, 2, 1$. This is the BS step. The BS procedure is implemented by first noting that ξ_{t+1} is the most recent value sampled for the hidden Markov chain at $t+1$ and, secondly, since ξ_t can take one of $e_1, \ldots, e_M$, we can compute the expression in Eq. (15) for each of these values. Sampling ξ_t from $q(\xi_t|\xi_{t+1}, \theta, y_{1:t})$ may be compared to multinomial sampling, provided that the probability of $\xi_i = e_i$, $i = 1, \ldots, M$, is known. Note that at each iteration step of the MCMC procedure, we only need a single run of the Forward Filter for generating multiple proposals using Backward Sampling.

A summary of our MTM algorithm is given in Algorithm 1.

Observe that the MTM algorithm reduces to the standard independent Metropolis–Hasting algorithm when $K = 1$. We also note that alternative weight functions other than the importance weights assumed in the MTM algorithm presented above could be used (e.g., see Craiu and Lemieux (2007)).

3.1.2. *Multiple-trial metropolized independent sampler*

As suggested by Liu (2002), when using independent proposal distributions, the generation of a set of reference points is not needed to design a possibly more efficient generalized Metropolis algorithm. Thus, we combine the FFBS proposals with the Liu (2002) metropolized independent sampler (MTMIS) and obtain Algorithm 2. The main advantage with respect to the MTM is that in MTMIS reference points are not used, obtaining thus a decrease of the computational complexity of the algorithm. It can be shown that the definition of acceptance probability at step v of Algorithm 2 guarantees the reversibility of the MTMIS chain.

Algorithm 2 Multiple-Trial Metropolized Independent Sampler (MTMIS)

---

i. Let $\xi_{1:T}^{(r-1)}$ be the value of the MTM at the $(r-1)$-th iteration.

ii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ containing K independent paths of the state variable drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$.

iii. Evaluate

$$W_k(\xi_{1:T,k}) = \frac{p(\xi_{1:T,k}|, \theta^{(r-1)}, y_{1:T})}{q(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T})}, \quad \forall\, k = 1, \ldots, K,$$

and define

$$W = \sum_{k=1}^{K} W_k(\xi_{1:T,k}).$$

iv. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,k})}{\sum_{k=1}^{K} W_k(\xi_{1:T,k})}, \quad \forall\, k = 1, \ldots, K.$$

v. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi_{1:T}^{(r)} = \begin{cases} \tilde{\xi}_{1:T}, & \text{if } u \leq \alpha_{MTMIS}, \\ \xi_{1:T}^{(r-1)}, & \text{otherwise,} \end{cases}$$

where,

$$\alpha_{MTMIS} = \min\left(1, \frac{W}{W - W_{k'}(\tilde{\xi}_{1:T}) + W_{k'}(\xi_{1:T}^{(r-1)})}\right),$$

and k' is the index of the selected trajectory among the trial set in step iv.

It is seen that the selection probability in Algorithm 2 is the same as in Algorithm 1. However, the acceptance probability in Algorithm 2 defer from that of Algorithm 1 by the composition of the denominator. The denominator of the acceptance

ratio in Algorithm 2 is evaluated as the sum of the numerator and the importance weight of the most recent hidden state less the selected sample from the trial set.

3.1.3. Multiple correlated-try Metropolis sampler

To further improve the mixing of the MTM chain, we propose the use of correlated proposals. There are various ways of introducing correlation among proposals (e.g. antithetic and stratified approaches). In this paper, we study the antithetic approach. The use of antithetic sampling in a Gibbs sampling context allows for a gain of efficiency. Pitt and Shephard (1996) propose a blocking method with an antithetic approach for non-Gaussian state space models, Holmes and Jasra (2009) propose a scheme for reducing the variance of estimates from the standard Metropolis-within-Gibbs sampler by introducing antithetic samples, while Bizjajeva and Olsson (2008) propose a forward filtering backward smoothing particle filter algorithm with antithetic proposals. Here, we follow Craiu and Lemieux (2007) who use antithetic proposals within a multi-point sampler and apply their idea to the context of discrete state space models. We propose a multiple correlated-try Metropolis sampler (MCTM) based on a combination of the FFBS and antithetic sampling techniques, and obtain Algorithm 3a. To the best of our knowledge, antithetic proposals of this kind have not been used in the context of Markov switching nonlinear state space models. The idea is to choose, at each step of the MCMC algorithm, a new hidden state trajectory from negatively correlated proposals instead of independent proposals.

Algorithm 3a Multiple Correlated-Try Metropolis Sampler (MCTM)

i. Let $\xi_{1:T}^{(r-1)}$ be the value of the MTM at the $(r-1)$-th iteration.
ii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ containing K correlated paths of the state variable drawn from the proposal distribution.
iii. Evaluate

$$W_1(\xi_{1:T,1}) = p(\xi_{1:T,1}|\theta^{(r-1)}, y_{1:T}),$$
$$W_k(\xi_{1:T,1:k}) = p(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T}) \prod_{i=1}^{k-1} q(\xi_{1:T,i}|\theta^{(r-1)}, y_{1:T}, \xi_{1:T,i:k}), \ \forall\ k = 2, \ldots, K.$$

iv. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,1:k})}{\sum_{k=1}^{K} W_k(\xi_{1:T,1:k})}, \quad \forall k = 1, \ldots, K.$$

v. Suppose $\tilde{\xi}_{1:T} = \xi_{1:T,l}$ is chosen in item (v) above, create a reference set $\{\xi^*_{1:T,1}, \xi^*_{1:T,2}, \ldots, \xi^*_{1:T,K}\}$ by letting

$$\xi^*_{1:T,j} = \xi_{1:T,l-1}, \quad \forall\ j = 1, \ldots, l-1,$$
$$\xi^*_{1:T,l} = \xi_{1:T}^{(r-1)},$$

and drawing $\xi^*_{1:T,j}$ for $j = l+1, \ldots, K$ from the proposal distribution.
vi. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi_{1:T}^{(r)} = \begin{cases} \tilde{\xi}_{1:T}, & \text{if } u \le \alpha_{MTCM}, \\ \xi_{1:T}^{(r-1)}, & \text{otherwise,} \end{cases}$$

where,

$$\alpha_{MTCM} = \min\left(1, \frac{\sum_{k=1}^{l} W_k(\xi_{1:T,1:k})}{\sum_{k=1}^{l} W_k(\xi^*_{1:T,1:k})}\right).$$

The simplest way to introduce the negative correlation between the trajectories generated with the FFBS algorithm is to use, at a given iteration r of the sampler and for the tth hidden state variable, a set of K uniform random numbers $U_{t,k}^{(r)}$, $k = 1, \ldots, K$, generated following the permuted displacement method (see Arvidsen and Johnsson (1982) and Craiu and Meng (2005)) given in Algorithm 3b. The uniform random numbers are then used within the BS procedure to generate correlated proposals.

For $K = 3$, Craiu and Meng (2005) show that the random numbers generated with the permuted displacement method are pairwise negatively associated (PNA). The definition of PNA given in the following is adopted from Craiu and Meng (2005).

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

Algorithm 3b Permuted displacement method

- Draw $r_1 \sim \mathcal{U}_{[0,1]}$.
- For $k = 2, \ldots, K-1$, set $r_k = \{2^{k-2} r_1 + 1/2\}$ where $\{x\}$ denotes the fractional part of x.
- Set $r_K = 1 - \{2^{K-2} r_1\}$.
- Pick at random $\sigma \in S_K$, where S_K is the set of all possible permutation of the integers $\{1, \ldots, K\}$.
- For $k = 1, \ldots, K$, set $U_k = r_{\sigma(k)}$.

Definition 1 (*Pairwise Negative Association*)**.** The random variables $\xi_{t,1}, \xi_{t,2}, \ldots, \xi_{t,K}$ are said to be pairwise negatively associated (PNA) if, for any nondecreasing functions f_1, f_2 and $(i, j) \in \{1, \ldots, K\}^2$ such that $i \neq j$

$$Cov(f_1(\xi_{t,i}), f_2(\xi_{t,j})) \leq 0,$$

whenever this covariance is well defined.

The proof for the case $K \geq 4$ is still an open issue. For this reason we consider in our algorithm $K \leq 3$. The presence of PNA in the case of $K \geq 4$ proposals depends upon the degree of uniformity in the filtering probability and the gain of efficiency should be proven computationally in each application.

We use the permuted sampler to generate $K = 2$ multi-move and correlated proposals in the backward sampling step of the FFBS. In order to show how the antithetic sampler works, we consider the case where the hidden Markov switching process has two states (i.e. $\xi_t = (\xi_{1t}, \xi_{2t})'$). For notational convenience let $\{q_t^{(r)}\}_{t=1,\ldots,T}$ be the sequence of filtered probabilities of being in state 1 at the rth iteration of the sampler.

Proposition 1. *Define backward antithetic samples $\xi_{t,1}$ and $\xi_{t,2}$ as follows*

$$\xi_{t,1} = \begin{pmatrix} \mathbb{I}_{\{U_t^{(r)} < q_t^{(r)}\}} \\ \mathbb{I}_{\{U_t^{(r)} \geq q_t^{(r)}\}} \end{pmatrix}, \qquad \xi_{t,2} = \begin{pmatrix} \mathbb{I}_{\{V_t^{(r)} < q_t^{(r)}\}} \\ \mathbb{I}_{\{V_t^{(r)} \geq q_t^{(r)}\}} \end{pmatrix},$$

where $V_t^{(r)} = 1 - U_t^{(r)}$ and $U_t^{(r)} \sim \mathcal{U}_{[0,1]}$. Then it is possible to show by following the rules of expectation that

$$Cov(\xi_{t,1}^{(r)}, \xi_{t,2}^{(r)}) = \begin{pmatrix} (2q_t^{(r)} - 1)\mathbb{I}_{\{q_t^{(r)} > \frac{1}{2}\}} - \left(q_t^{(r)}\right)^2 & \left(q_t^{(r)}(1 - q_t^{(r)})\right)^2 \\ \left(q_t^{(r)}(1 - q_t^{(r)})\right)^2 & (1 - 2q_t^{(r)})\mathbb{I}_{\{q_t^{(r)} < \frac{1}{2}\}} - \left(1 - q_t^{(r)}\right)^2 \end{pmatrix}.$$

Proof. The result follows from the application of expectation properties.

Using the expected value of the square of the Euclidean distance, $d(\xi_{t,1}, \xi_{t,2})$, between these two antithetic samples, that is

$$E[d^2(\xi_{t,1}, \xi_{t,2})] = 2 - 2\left((2q_t^{(r)} - 1)\mathbb{I}_{\{q_t^{(r)} > \frac{1}{2}\}} + (1 - 2q_t^{(r)})\mathbb{I}_{\{q_t^{(r)} < \frac{1}{2}\}}\right), \tag{16}$$

it is possible to verify that extremely antithetic proposals are obtained when the average distance is optimal. From Eq. (16), extreme antithetic samples are obtained when $q_t^{(r)}$ is equal to 0.5, which can be easily found in applications where regimes exhibit a similar persistence level.

3.2. Auxiliary models for defining the proposal distribution

In order to construct efficient proposal distributions for the state variables, first, we approximate the MS-GARCH model by eliminating the problem of path dependence; and second we apply FFBS to the auxiliary model thus obtained. A possible way of circumventing the path dependence problem inherent in the MS-GARCH model is to replace the lagged conditional variance appearing in the definition of the GARCH model with a proxy. In line with this, we develop a class of MS-GARCH models based on a collapsing procedure:

$$y_t = \mu_t(y_{1:t-1}, \xi_t'\theta_\mu) + \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0, 1), \tag{17}$$

$$\sigma_t^2 \approx \tilde{\sigma}_t^2 = (\xi_t'\gamma) + (\xi_{t_1}'\alpha)\epsilon_{(X)t-1}^2 + (\xi_{t_2}'\beta)\sigma_{(X)t-1}^2, \tag{18}$$

where $t_1, t_2 \leq t$,

$$\sigma_{(X)t-1}^2 = V(E[y_{t-1}|\mathbb{1}_{2t}]|\mathbb{1}_{1t}) + E[V(y_{t-1}|\mathbb{1}_{3t})|\mathbb{1}_{4t}],$$

$$\epsilon_{(X)t-1} = y_{t-1} - E[\mu_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\mu)|\mathbb{1}_{5t}],$$

and $\mathcal{I}_{1t}$, $\mathcal{I}_{2t}$, $\mathcal{I}_{3t}$, $\mathcal{I}_{4t}$, and $\mathcal{I}_{5t}$ are the conditioning sets at time t. The GARCH parameters and the transition probabilities remain as defined in Section 2.1. Alternative specifications of $\epsilon_{(X)t-1}$ and $\sigma^2_{(X)t-1}$ amount to different MS-GARCH auxiliary models. In the following, the label X is replaced with one of B, G, D, K, KL, representing different approximations. Define $\mathcal{I}^{\xi}_{t-1} = \{\xi_1, \ldots, \xi_{t-1}\}$ and $\mathcal{I}^{Y}_{t-1} = \{y_{1:t-2}\}$. Then

1. Gray's approximation (model G): Gray (1996) model is obtained by setting $t_1 = t_2 = t$, $\mathcal{I}_{1t} = \mathcal{I}^{Y}_{t-1}$, $\mathcal{I}_{2t} = \mathcal{I}^{Y}_{t-1} \cup \mathcal{I}^{\xi}_{t}$, $\mathcal{I}_{3t} = \mathcal{I}^{Y}_{t-1} \cup \mathcal{I}^{\xi}_{t}$, and $\mathcal{I}_{4t} = \mathcal{I}_{5t} = \mathcal{I}^{Y}_{t-1}$, i.e.

$$\tilde{\sigma}^2_t = (\xi'_t\gamma) + (\xi'_t\alpha)\epsilon^2_{(G)t-1} + (\xi'_t\beta)\sigma^2_{(G)t-1},$$

with

$$\epsilon_{(G)t-1} = y_{t-1} - \mu_{(G)t-1},$$
$$\mu_{(G)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}],$$
$$\sigma^2_{(G)t-1} = V(\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}) + E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-2}].$$

2. Dueker's approximation (model D): Dueker (1997) model may be obtained by setting $t_1 = t_2 = t-1$, $\mathcal{I}_{1t} = \mathcal{I}^{Y}_{t} \cup \mathcal{I}^{\xi}_{t}$, $\mathcal{I}_{2t} = \mathcal{I}^{Y}_{t}$, $\mathcal{I}_{3t} = \mathcal{I}^{Y}_{t-1} \cup \mathcal{I}^{\xi}_{t}$, and $\mathcal{I}_{4t} = \mathcal{I}_{5t} = \mathcal{I}^{Y}_{t}$, i.e.

$$\tilde{\sigma}^2_t = (\xi'_t\gamma) + (\xi'_{t-1}\alpha)\epsilon^2_{(D)t-1} + (\xi'_{t-1}\beta)\sigma^2_{(D)t-1}(\xi'_{t-1}),$$

with

$$\epsilon_{(D)t-1} = y_{t-1} - \mu_{(D)t-1},$$
$$\mu_{(D)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}],$$
$$\sigma^2_{(D)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-2}\theta_\sigma)|y_{1:t-1}].$$

3. Klaassen's approximation (model K): Klaassen (2002) MS-GARCH(1, 1) model is obtained by setting $t_1 = t_2 = t$, $\mathcal{I}_{1t} = \mathcal{I}^{Y}_{t} \cup \mathcal{I}^{\xi}_{t}$, $\mathcal{I}_{2t} = \mathcal{I}^{Y}_{t-1}$, $\mathcal{I}_{3t} = \mathcal{I}^{Y}_{t-1} \cup \mathcal{I}^{\xi}_{t+1}$, and $\mathcal{I}_{4t} = \mathcal{I}_{5t} = \mathcal{I}^{Y}_{t} \cup \{\xi_t\}$, i.e.

$$\tilde{\sigma}^2_t = (\xi'_t\gamma) + (\xi'_t\alpha)\epsilon^2_{(K)t-1} + (\xi'_t\beta)\sigma^2_{(K)t-1}(\xi'_t),$$

with

$$\epsilon_{(K)t-1} = y_{t-1} - \xi'_t\mu_{(K)t-1},$$
$$\mu_{i,(K)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}, \xi_t = e_i],$$
$$\sigma^2_{i,(K)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-1}, \xi_t = e_i],$$

where $\mu_{(K)t-1} = (\mu_{1,(K)t-1}, \ldots, \mu_{M,(K)t-1})$.

In addition, Eqs. (17) and (18) also give many new MS-GARCH models not yet considered in the literature. For example, we identify the following new MS-GARCH models and tag them as the Basic and the Simplified Klaassen approximation.

1. The Basic approximation (model B): this is obtained by setting $t_1 = t_2 = t$, $\mathcal{I}_{1t} = \mathcal{I}^{Y}_{t} \cup \mathcal{I}^{\xi}_{t}$, $\mathcal{I}_{2t} = \mathcal{I}^{Y}_{t}$, $\mathcal{I}_{3t} = \mathcal{I}^{Y}_{t-1} \cup \mathcal{I}^{\xi}_{t}$, and $\mathcal{I}_{4t} = \mathcal{I}_{5t} = \mathcal{I}^{Y}_{t-1}$, i.e.

$$\tilde{\sigma}^2_t = (\xi'_t\gamma) + (\xi'_t\alpha)\epsilon^2_{(B)t-1} + (\xi'_t\beta)\sigma^2_{(B)t-1},$$

with

$$\epsilon_{(B)t-1} = y_{t-1} - \mu_{(B)t-1},$$
$$\mu_{(B)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}] = E[y_{t-1}|y_{1:t-2}]$$
$$= \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu)q(\xi_{t-1} = e_m|y_{1:t-2}),$$
$$\sigma^2_{(B)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-2}] = E[\epsilon^2_{t-1}|y_{1:t-2}] = V(\epsilon_{t-1}|y_{1:t-2})$$
$$= \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)q(\xi_{t-1} = e_m|y_{1:t-2}).$$

If $\mu_t = 0\ \forall\, t$, then this approximation is equivalent to Gray's approximation. In this approximation scheme $\mu_{(B)t-1}$ and $\sigma^2_{(B)t-1}$ are functions of $y_{1:t-2}$ and the information coming from y_{t-1} is lost. In line with Gray's approach, $\sigma^2_{(B)t-1}$ is equal to the variance of the conditional density of ϵ_t. For practical implementation of this approximation, given $q(\xi_{t-1} = e_m|y_{1:t-2})$ for $m = 1, \ldots, M$, $\mu_{(B)t-1}$ can easily be computed while $\sigma^2_{(B)t-1}$ can be computed recursively since $\sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)$ depends on $\sigma^2_{(B)t-2}$. This auxiliary model uses the least information ($y_{1:t-2}$) among the auxiliary models under consideration. Hence, we tag it as the basic model approximation.

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

2. The simplified Klaassen approximation (model SK) is obtained by letting $t_1 = t_2 = t$, $\mathcal{I}_{1t} = \mathcal{I}_t^Y \cup \mathcal{I}_t^\xi$, $\mathcal{I}_{2t} = \mathcal{I}_t^Y$, $\mathcal{I}_{3t} = \mathcal{I}_{t-1}^Y \cup \mathcal{I}_t^\xi$, and $\mathcal{I}_{4t} = \mathcal{I}_{5t} = \mathcal{I}_t^Y$, i.e.

$$\tilde{\sigma}_t^2 = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon_{(SK)t-1}^2 + (\xi_t'\beta)\sigma_{(SK)t-1}^2,$$

with

$$\epsilon_{(SK)t-1} = y_{t-1} - \mu_{(SK)t-1},$$

$$\mu_{(SK)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\mu)|y_{1:t-1}] = \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e_m'\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-1}),$$

$$\sigma_{(SK)t-1}^2 = E[\sigma_{t-1}^2(y_{1:t-2}, \xi_{t-1}'\theta_\sigma)|y_{1:t-1}] = \sum_{m=1}^{M} \sigma_{t-1}^2(y_{1:t-2}, e_m'\theta_\sigma) q(\xi_{t-1} = e_m|y_{1:t-1}).$$

This approximation is similar to Dueker's approximation (Model D). As opposed to Dueker's approximation, we assume that σ_{t-1}^2 is a function of $(y_{1:t-2}, \xi_{t-1})$. On the other hand, Klaassen's approximation reduces to the simplified Klaassen approximation model by eliminating ξ_t' from the conditioning set.

3.3. Sampling θ

Sampling θ from its full conditional distribution will be done by separating the parameters of the transition matrix from the GARCH parameters, accordingly. We assume that the parameters of the transition probabilities are independent of the GARCH parameters.

With regard to the transition probability parameters, θ_π, their posterior distribution is given by the product of independent Dirichlet distributions

$$f(\theta_\pi|\xi_{1:T}, \theta_\mu, \theta_\sigma, y_{1:T}) = \prod_{m=1}^{M} \text{Dirichlet}(n_{1m} + \eta_{1m}, \ldots, n_{Mm} + \eta_{Mm}), \tag{19}$$

where

$$n_{ij} = \sum_{t=1}^{T} \xi_{jt-1}\xi_{it}.$$

This distribution is standard and can be simulate from directly.

Given a prior density $f(\theta_\mu, \theta_\sigma)$, the posterior density of the GARCH parameters, $(\theta_\mu, \theta_\sigma)$, can be expressed as

$$f(\theta_\mu, \theta_\sigma|\xi_{1:T}, \theta_\pi, y_{1:T}) \propto f(\theta_\mu, \theta_\sigma) \prod_{t=1}^{T} \mathcal{N}(\mu_t(y_{1:t-1}, \xi_t'\theta_\mu), \sigma_t^2(y_{1:t-1}, \xi_t'\theta_\sigma)). \tag{20}$$

For this step of the Gibbs sampler, we apply an adaptive Metropolis–Hastings (MH) algorithm since the full conditional distribution is known to be non-standard. Details can be found, as required, in Section A of the Supplementary Material (see Appendix A).

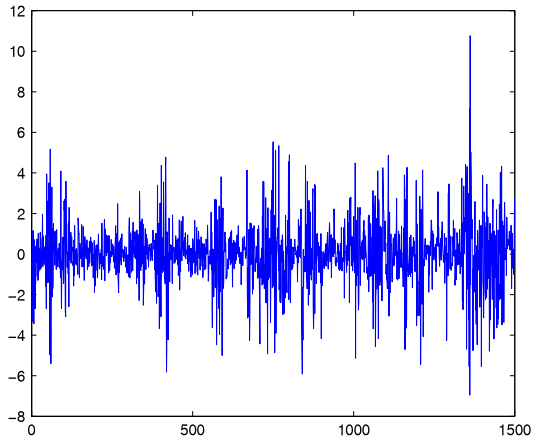
4. Monte Carlo experiments on simulated data sets

In this section, we present two simulation studies to evaluate the performance of our multi-move and multi-point algorithms. The first study compares all of the multi-move and multi-point algorithms with the particle MCMC (PMCMC) approach proposed by Bauwens et al. (2014). The second deals with the choice of the optimal blocking strategy and the number of proposals. Note that conditions for blocking to improve and speed up convergence have been studied by Roberts and Sahu (1997) for Gaussian target distributions (see also Liu and Wong (1994), and Besag et al. (1995)). However, in practice the optimal blocking strategy needs to be studied for the specific problem and algorithm used (see Fiorentini et al. (2012) for illustration). Thus, in the second set of experiments we study the performance of one of the proposed multi-move and multi-point algorithms, when one of the regimes is highly persistent (i.e. $\alpha + \beta$, is between 0.9 and 1).

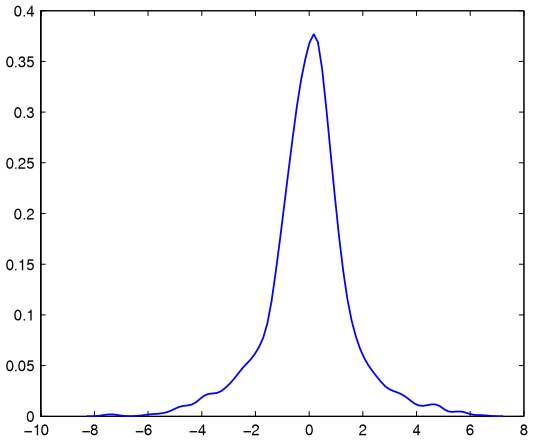
4.1. A comparison of algorithms

In the first simulation exercise, we generate a time series of length 1500 from the data generating process (DGP) corresponding to the model defined by Eqs. (3) and (4) for two regimes ($M = 2$), including their time invariant transition probabilities and switching conditional mean and variance. We set $(\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.30, 2.00)$, $(\alpha_1, \alpha_2) = (0.35, 0.10)$, $(\beta_1, \beta_2) = (0.20, 0.60)$, and $\pi_{11} = 0.98$, $\pi_{22} = 0.96$. This parameter setting corresponds to the

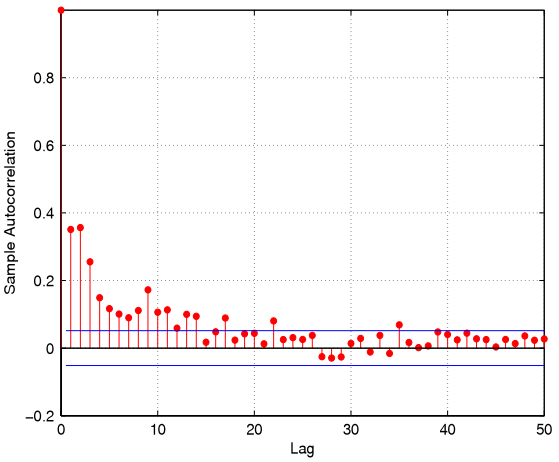
Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011



(a) Time series.



(b) Density approximation.



(c) Sample autocorrelation of squared returns.

Fig. 1. Simulated data for a MS-GARCH model with parameter setting: $(\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.30, 2.00)$, $(\alpha_1, \alpha_2) = (0.35, 0.10)$, $(\beta_1, \beta_2) = (0.20, 0.60)$, and $\pi_{11} = 0.98, \pi_{22} = 0.96$.

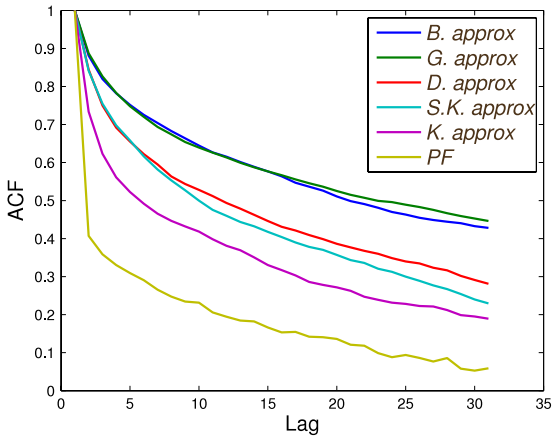
one in Bauwens et al. (2010) for a similar Monte Carlo exercise. A relatively higher and more persistent conditional variance, as compared to the first regime GARCH equation, is implied by the second regime GARCH equation. Also, the probability of staying in each regime is close to one. A typical series simulated from this DGP exhibits volatility clusters (see Fig. 1(a)). The kernel estimate of the unconditional density has heavy tails (see Fig. 1(b)), and the excess kurtosis is estimated to be 3.57. The autocorrelation function (ACF) of the square of the same series (Fig. 1(c)) is significant and this calls for the use of autoregressive volatility models.

For each sampling algorithm described in Section 3.1 and the auxiliary model presented in Section 3.2, we perform 10 000 Gibbs iterations after convergence. We assess the convergence by testing for the equality of means between the first 10% and the last 50% samples of the converging chain (Geweke, 1992). It is well known that the Gibbs sampler produces drawings with very high positive serial correlation and the results for consecutive samples may yield poor approximations to the posterior quantities of interest. In view of this, we consider every 10th draw after convergence of the Gibbs chain when evaluating the results presented in this section. The MCMC exercise is carried out by setting the initial state trajectory and parameter values of the algorithm to the maximum likelihood estimates of the MS-GARCH model based on the Basic model approximation (see Section 3.1). The hyperparameters, ν_{ij} for $i, j = 1, 2$, of the prior distributions of the transition probabilities are set equal to 1. The support for the other parameters are defined to obey their respective constraints. The case of two trials, $(K = 2)$, is considered within the different multi-point sampling strategies discussed earlier.

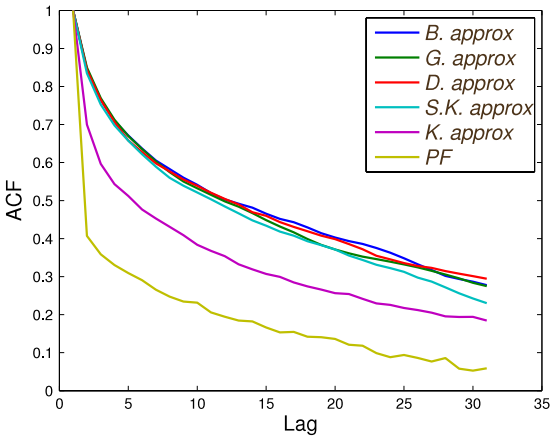
Using the first 10 000 draws of the MCMC algorithm, we compare the efficiency of the different multi-move hidden state sampling algorithms against the PMCMC approach by computing the autocorrelations from lag 1 to 30 for the number of observations allocated to the first regime over the MCMC iterations. In Figs. 2 and 3, we plot the autocorrelations for each of the multi-move multi-point sampling algorithms and the PMCMC scheme.

From Fig. 2, it may be deduced that under each multi-point sampling scheme the PMCMC algorithm produces the least autocorrelations while the Klaassen (2002) approximation produces the least autocorrelations among the auxiliary MS-

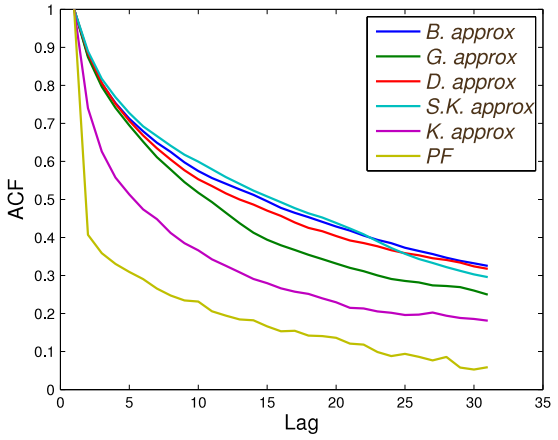
Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011



(a) Multiple-try Metropolis (MTM).



(b) Multiple-trial metropolized independent sampler (MTMIS).



(c) Multiple correlated-try Metropolis (MCTM).

Fig. 2. Autocorrelation of the number of observations of being in state 1, for each sampler (different plots) and approximation methods (different lines in each plot). Note: PF., B. approx, G. approx, D. approx, S. K. approx and K. approx respectively represent Particle MCMC, Basic model approximation, Gray's approximation, Dueker's approximation, simplified Klaassen's approximation and Klaassen's approximation.

GARCH models considered. Except in the case of the MTM sampling strategy (see Fig. 2(a)), there is no clear difference among the autocorrelations produced by other approximations in each of the sampling scheme. On the other hand, from Fig. 3 we observe that MTMIS consistently performs at least as well as other sampling schemes. The efficiency of the various multi-move and multi-point samplers are further assessed by computing the inefficiency factor and the relative inefficiency of the multi-point algorithms.

Let $\#(\xi_{1:T})^{(1)}, \ldots, \#(\xi_{1:T})^{(G)}$, denote a sample from the posterior distribution of a random variable $\#(\xi_{1:T})$ representing the number of observations of being in state 1. Then the inefficiency factor (IF) is

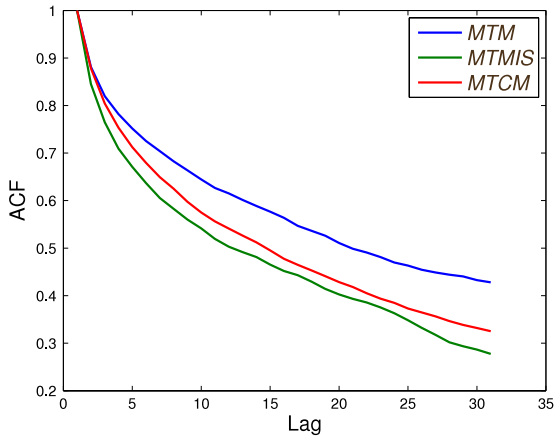
$$IF = 1 + 2\sum_{l=1}^{L} w_l \rho_l, \tag{21}$$

where $\rho_l, l = 1, 2, \ldots, L$ is the autocorrelation function of $\#(\xi_{1:T})^{(1)}, \ldots, \#(\xi_{1:T})^{(G)}$ at lag l and w_l is the associated weight (see Robert and Casella (2010) for details). If the samples are independent, then $IF = 1$. We define the relative inefficiency (RI) factor of two competing algorithms A and B with inefficiency factors given by IF_A and IF_B, respectively, as

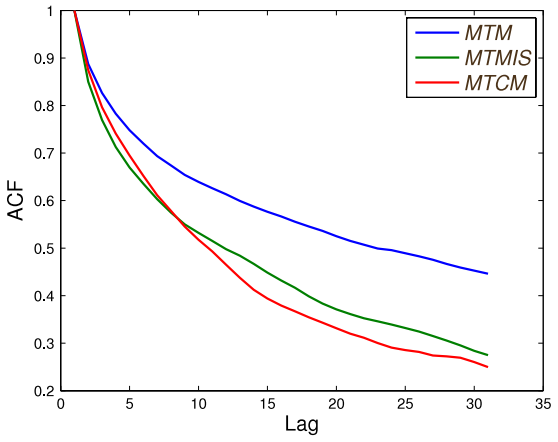
$$RI = \frac{Time_A}{Time_B} \times \frac{IF_A}{IF_B}, \tag{22}$$

where $Time_A$ and $Time_B$ are the computing times of each algorithm. RI measures the factor by which the run-time of algorithm A must be increased to achieve algorithm B's precision; values greater than one suggest that algorithm B is more efficient. In Table 1, we present the IF and report the RI for various multi-move and multi-point algorithms relative to the PMCMC sampling technique. The number of lags over which we calculate both the IF and RI is set equal to $L = 500$. The inefficiency factor shows that MTMIS generally performs best among the multi-point sampling techniques, while Klaassen (2002) is the

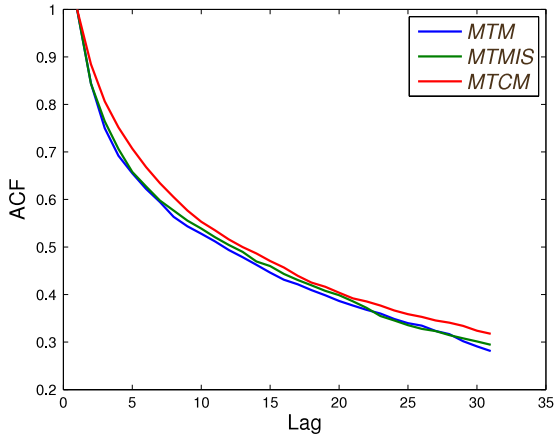
Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011



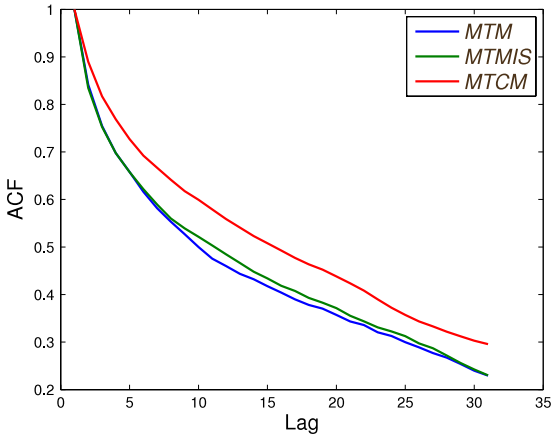
(a) Basic model approximation.



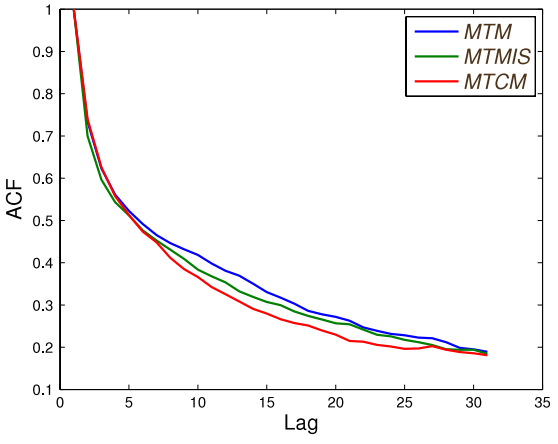
(b) Gray (1996) approximation.



(c) Dueker (1997) approximation.



(d) Simplified Klaassen (2002) approximation.



(e) Klaassen (2002) approximation.

Fig. 3. Autocorrelation of the number of observations in regime 1, for each approximation method (different plots) and samplers (different lines in each plot).

best among the approximation methods (see Table 1). We equally observe that our multi-move and multi-point algorithms are more efficient than the PMCMC sampling technique for the state variable (see columns *RI* in Table 1). Similarly, we observe from Table 1 that MTMIS under Klaassen's approximation is the most efficient combination. It does not come too much as a surprise that there is no appreciable difference in the efficiency of the MTCM over standard MTM because of the small number of proposals used (see Craiu and Lemieux (2007) for a discussion).

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

Table 1
Inefficiency (IF) and relative inefficiency (RI) factors.

	PMCMC	MTM		MTMIS		MTCM	
	IF	IF	RI	IF	RI	IF	RI
B. approx.	3.30	7.64	0.13	6.55	0.11	6.91	0.10
G. approx.	3.30	7.94	0.14	6.59	0.11	6.47	0.10
D. approx.	3.30	6.28	0.11	6.56	0.11	6.86	0.10
S. K. approx.	3.30	6.37	0.11	6.22	0.10	6.22	0.10
K. approx.	3.30	5.29	0.09	5.19	0.09	5.19	0.09

Notes: IF and RI as defined in Eq. (21) and Eq. (22) respectively are computed using 10 000 draws; B. approx, G. approx, D. approx, S. K. approx and K. approx respectively represent Basic model approximation, Gray's approximation, Dueker's approximation, simplified Klaassen's approximation and Klaassen's approximation.

In Table 2, we report the posterior means and standard deviations of the parameters and the transition probabilities of the MS-GARCH using the various sampling schemes and approximation methods. With the exception of a few cases, the estimated posterior means of the parameters have more values within one posterior standard deviation away from their true values. We also quantify the performance of these estimates by calculating the mean squared error (MSE) of the posterior means of parameters relative to the true parameters i.e.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2, \tag{23}$$

where n is the number of parameters, and $\hat{\theta}_i$ is the estimate of the ith element, θ_i, of the DGP parameter set.

The results in Table 2 show that some of the parameter estimates are a bit off their respective true values. A possible explanation for this bias could be related to the variance scaling factor of the proposal distribution for sampling the GARCH parameters, which needs to be fine-tuned in order to obtain estimates closer to the true parameters. Another explanation may be related to the choice of identification constraints. In our simulation exercise we considered a simple ordering of the intercept ($\gamma_1 < \cdots < \gamma_M$) of the MS-GARCH model in order to identify the regimes. This approach, as illustrated in Marin et al. (2005) and Jasra et al. (2005), may have some consequences (such as large bias) on the resulting inference. In view of this, we leave for further research the study of alternative identification constraints or of different solutions to the identification problem, such as the random permutation MCMC sampler of Frühwirth-Schnatter (2006) and the re-labeling algorithm of Jasra et al. (2005) on the parameter estimates. Perhaps a different choice of the identification constraint may bring about a reduction in bias. The large standard deviation observed for some of the parameters in the second regime may be attributed to the fact that the regime is less frequently visited. Estimates of our multi-point and multi-move sampling schemes are also observed to be very similar. This result is not too surprising because the different approximations are based on the same fundamental, while the different sampling strategies are meant to improve the efficiency of the MCMC algorithm. Furthermore, the assumption of low persistence ($\alpha_m + \beta_m \ll 1$ for $m = 1, 2$) for all the regimes may also have contributed to obtaining close parameter estimates. As with the output of maximization algorithms, MCMC parameter estimates may change when they operate close to the boundaries of the constrained domain (i.e. when MS-GARCH is strongly persistent in each regime). Nevertheless, unlike the PMCMC state sampling algorithm that is prone to severe degeneracy problem (as experienced in our exercise), which in turn may lead to poor parameter estimates (see Table 2), our multi-point and multi-move state sampling schemes are capable of avoiding such a scenario. The potentials of our multi-point sampling schemes are further made clear by their low MSEs (see Table 2).

Fig. B.1–B.3 in the Supplementary Material (see Appendix A) reports the posterior densities of the parameters obtained via the different sampling strategies. The posterior densities are unimodal, thus ruling out the presence of label switching. From these figures, we observe that the PMCMC (blue line) algorithm produces parameter density estimates centered away from their corresponding true values. The multi-point and multi-move algorithms, on the other hand, produce parameter posterior densities concentrated around the true values. As noted for the parameter estimates, there is no clear difference among the density plots obtained using the multi-point and multi-move schemes.

It may be argued, as it is often the case, that the starting values for our MCMC algorithm may influence the parameter estimates and the results reported in Table 2 may have occurred by chance. To rule out this criticism, the previous experiment is repeated 50 times, thus generating 50 samples of size 1500 of the same DGP and repeating the MCMC estimation for each sample using the same set of starting values. Fig. C.1–C.3 in the Supplementary Material (see Appendix A) displays a summary of each set of parameter estimates. A look at these figures show that regime 2 parameters are more difficult to estimate, with the PMCMC algorithm producing the largest deviation from their true values.

Based on these 50 replications, we compute the average burn-in period required to achieve convergence following Geweke's convergence criterion. Table 3 shows that, on average, the Klaassen (2002) approximation seems to converge the fastest among the different approximation methods, while MTMIS offers the shortest burn-in period among the multi-point sampling schemes.

The performance of our multi-move, multi-point algorithms relative to the PMCMC strategy is further examined by computing the percentage of correctly specified regimes. To do this, for each of the 50 replications, we calculate the average of the Gibbs output on the state variables and then assign mean states greater than one-half to regime 2 (regime 1 otherwise).

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

Table 2
Estimated posterior mean and standard deviation of MS-GARCH parameters using various state proposal distributions and multi-move algorithms.

	DGP values	PMCMC	Multi-move B. approx.			Multi-move G. approx.			Multi-move D. approx.			Multi-move S. K. approx.			Multi-move K. approx.		
			MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM
π_{11}	0.980	0.985	0.987	0.987	0.987	0.986	0.987	0.987	0.986	0.987	0.987	0.987	0.987	0.987	0.987	0.986	0.986
		(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
π_{22}	0.960	0.963	0.961	0.960	0.961	0.961	0.960	0.962	0.961	0.961	0.961	0.961	0.961	0.962	0.961	0.960	0.960
		(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
μ_1	0.060	0.051	0.045	0.045	0.044	0.045	0.044	0.044	0.045	0.044	0.045	0.044	0.044	0.044	0.045	0.045	0.045
		(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
μ_2	−0.090	−0.085	−0.151	−0.155	−0.150	−0.151	−0.153	−0.152	−0.155	−0.156	−0.155	−0.152	−0.153	−0.154	−0.154	−0.156	−0.154
		(0.123)	(0.094)	(0.094)	(0.095)	(0.092)	(0.092)	(0.090)	(0.094)	(0.093)	(0.095)	(0.094)	(0.094)	(0.094)	(0.092)	(0.097)	(0.093)
γ_1	0.300	0.358	0.349	0.348	0.351	0.350	0.347	0.348	0.354	0.347	0.349	0.349	0.348	0.349	0.350	0.347	0.349
		(0.044)	(0.039)	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.039)	(0.039)	(0.038)	(0.041)	(0.040)	(0.039)	(0.038)	(0.040)	(0.039)
γ_2	2.000	4.240	2.770	2.883	2.851	2.923	2.877	3.005	2.855	2.993	2.901	2.992	2.959	2.895	2.907	2.982	2.853
		(0.960)	(1.110)	(1.132)	(1.155)	(1.138)	(1.078)	(1.147)	(1.137)	(1.099)	(1.127)	(1.160)	(1.070)	(1.088)	(1.156)	(1.125)	(1.072)
α_1	0.350	0.342	0.411	0.405	0.410	0.402	0.401	0.415	0.406	0.404	0.404	0.402	0.405	0.404	0.404	0.405	0.400
		(0.061)	(0.066)	(0.069)	(0.068)	(0.067)	(0.067)	(0.067)	(0.070)	(0.067)	(0.064)	(0.064)	(0.064)	(0.065)	(0.065)	(0.067)	(0.066)
α_2	0.100	0.166	0.111	0.108	0.109	0.111	0.112	0.112	0.109	0.110	0.110	0.107	0.111	0.111	0.110	0.111	0.112
		(0.079)	(0.055)	(0.057)	(0.055)	(0.055)	(0.055)	(0.058)	(0.057)	(0.056)	(0.057)	(0.053)	(0.058)	(0.054)	(0.057)	(0.057)	(0.056)
β_1	0.200	0.107	0.099	0.103	0.098	0.103	0.106	0.100	0.094	0.105	0.098	0.103	0.106	0.102	0.101	0.105	0.104
		(0.064)	(0.057)	(0.060)	(0.059)	(0.059)	(0.062)	(0.059)	(0.055)	(0.058)	(0.059)	(0.061)	(0.059)	(0.059)	(0.059)	(0.061)	(0.059)
β_2	0.600	0.218	0.473	0.460	0.462	0.448	0.457	0.439	0.463	0.440	0.453	0.444	0.444	0.454	0.457	0.442	0.459
		(0.131)	(0.177)	(0.176)	(0.183)	(0.182)	(0.166)	(0.178)	(0.182)	(0.176)	(0.177)	(0.181)	(0.170)	(0.173)	(0.181)	(0.174)	(0.176)
MSE		0.519	0.063	0.082	0.076	0.089	0.081	0.106	0.077	0.103	0.086	0.103	0.096	0.084	0.086	0.101	0.077

Notes: Columns 3–18: Posterior means and standard deviations (in parentheses) for the parameters of the MS-GARCH model. Results are based on a single sample of 1500 observations from the DGP defined by Eqs. (3) and (4); MSE is computed using Eq. (23).

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

Table 3
Estimated burn-in period.

	PMCMC	Multi-move algorithms		
		MTM	MTMIS	MCTM
B. approx.	1245	3550	2259	3648
G. approx.	1245	1836	2090	3750
D. approx.	1245	2579	2090	3667
S. K. approx.	1245	5209	3113	5587
K. approx.	1245	1208	2366	1816

Notes: The burn-in period is evaluated based on 10 000 Gibbs iterations after convergence according to Geweke's diagnostic. The estimated burn-in period is computed by taking the average burn-in period over 50 replications of the experiment.

We find out that the PMCMC technique is able to classify, on average, 90% of the data correctly, while the multi-move, multi-point samplers classified between 93% and 96% of the data correctly. The acceptance rate of the multi-move, multi-point proposals varies between 10% and 35%, with the highest arising from MTMIS sampling schemes characterized by a proposal distribution constructed using Klaassen's approximation. While, on average the acceptance rate of the mean parameters and the GARCH parameters are found to be about 54% and 30%, respectively.

Using our proposed multi-point sampling scheme, we observe a substantial computational time reduction, as expected, when compared with the PMCMC scheme. We also observe that the computational time required by our multi-point schemes are not different from one another. For one to observe an appreciable difference in the computational times among our multi-point sampling schemes, it will be required to increase the number of trial points. Finally, comparing our MCMC algorithm in terms of its coding cost, the basic model approximation is the easiest approximation to code, while Klaassen's approximation is the most complex and challenging approximation to code. MTMIS, on the other hand, is the simplest multi-point sampling scheme among the set under consideration to code.

4.2. Blocking strategies and number of proposals

The second study focuses on the performance of MTMIS in combination with the simplified Klaassen's approximation when one of the regimes is highly persistent (i.e. $\alpha + \beta$ is between 0.9 and 1). Especially for large values of β, the acceptance rate drops rapidly, making it necessary to search for other methods of improving the inefficiencies of the sampling procedure. In view of this, we investigate the effect of blocking and/or the number of trial points on the acceptance rate. We implement a blocking scheme where, for each run of the MCMC iteration, the block length is randomly chosen from a pre-determined set (i.e. $\{500, 750, 1500\}$). Two different numbers of trials are considered; namely $K = 2$ and 10. Hence, four different combinations of the blocking scheme and multi-points were considered: full block with two-trial points (FB-2 trials), full block with ten-trial points (FB-10 trials), random block scheme with two-trial points (RB-2 trials) and random block scheme with ten-trial points (RB-10 trials). In the second simulation exercise, we generate a time series of length 1500 and set $(\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.40, 0.60)$, $(\alpha_1, \alpha_2) = (0.20, 0.05)$, $(\beta_1, \beta_2) = (0.40, 0.90)$, and $\pi_{11} = 0.99$, $\pi_{22} = 0.99$. Using 50 replications of the experiments, an average acceptance rate of 5%, 3%, 9%, and 21% corresponding, respectively, to FB-2 trials, FB-10 trials, RB-2 trials and RB-10 trials, was obtained. Furthermore, the results of our parameter estimates indicate that the least bias is obtained for the RB-10 trials, while FB-2 trials produces the largest bias. These results suggest that without blocking, the MCMC chain easily gets trapped by local modes and thus results in a very inefficient sampling algorithm. The above experiment also demonstrates that both blocking and the multi-trial points are important to allow iterates to jump between different modes of the posterior density and to reduce correlations among iterates. Another interesting observation from our experiments is the noticeable increase in the computational time when estimates of parameters close to the boundary of the parameter space are required.

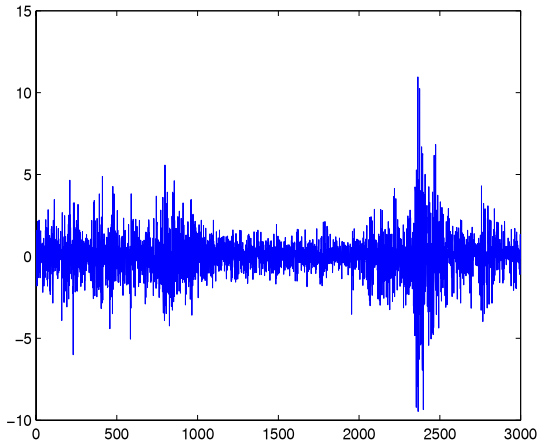
In summary, subject to the various performance criterion discussed above, Klaassen's approximation consistently performs at least as well as the other MS-GARCH approximations that were used in constructing proposal distributions. However, the complexity of coding Klaassen's model limits its attractiveness. Subject to this, and based on the fact that the parameter estimates obtained using the various multi-point, multi-move algorithms are not significantly different from each other, we suggest the use of the simplified Klaassen's approximation that is closely related to Klaassen's approximation and whose implementation is not as complex as Klaassen's approximation. Among the multi-point sampling schemes we suggest that the MTMIS should be used for practical purposes since it is the most efficient and requires the least coding cost. To improve on the quality of our estimates, blocking and multi-trial points are recommended.

5. Empirical application: the S&P500 daily returns

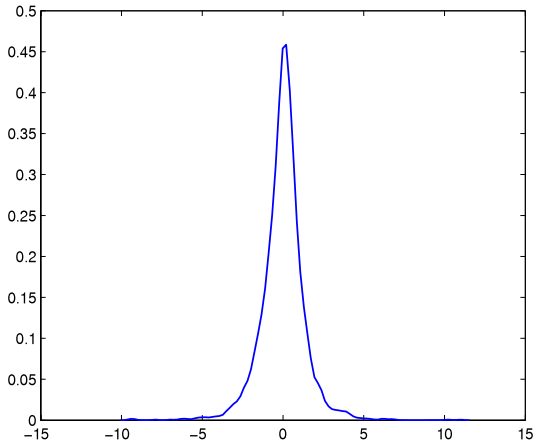
We use our proposed Bayesian estimation procedure to fit an MS-GARCH model to the S&P500 daily percentage returns from 20/05/1999 to 25/04/2011[1] (3000 observations). The Bauwens et al. (2014) estimates will serve as a benchmark

[1] We would like to thank Arnaud Dufays for making the data available on https://sites.google.com/site/websiteofarnauddufays/.

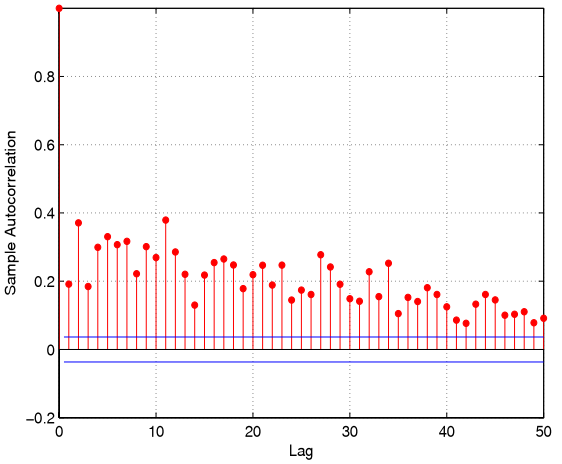
Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011



(a) Time series.



(b) Density approximation.



(c) Sample autocorrelation of squared returns.

Fig. 4. Graphs for S&P 500 daily returns from 20/05/1999 to 25/04/2011.

Table 4
Descriptive statistics for S&P 500 daily returns.

Min.	Max.	Mean	Std.	Skewness	Kurtosis
−9.470	10.960	−0.00022	1.353	−0.116	10.546

for evaluating the parameter estimates obtained using the MTMIS algorithm coupled with the simplified Klaassen's approximation and random blocking scheme. Fig. 4 displays the returns' sample path and kernel density estimate, and the autocorrelation of the squared returns. As one would expect from a typical financial time series, it exhibits strong persistence in the squared returns (see Fig. 4), slightly negative skewness and large excess kurtosis (see Table 4). These features call for the use of an MS-GARCH model.

Fitting a full MS-GARCH model to empirical data can lead to parameter estimates close to the boundary of the parameter space and often results in slow convergence of the MCMC chain. Based on this, a common practice in the estimation of MS-GARCH models with empirical data is to impose some restriction on the parameters. See Bauwens et al. (2010) and Augustyniak (2013) for illustrations. In this paper, two different specifications of the MS-GARCH model are considered. The first is tagged as the unconstrained model. In this case, we provide estimates for all the parameters of the MS-GARCH model. The second corresponds to specifications as laid out by Bauwens et al. (2014). In this case, we assume that $\mu_1 = \mu_2 = 0$. This case is tagged as the constrained model. In Table 5, we report the posterior means and standard deviations (in brackets) for both the constrained and unconstrained models. These parameter estimates are based on 2000 MCMC samples (i.e. we perform 20 000 Gibbs iterations after convergence according to Geweke's diagnostic and consider every 10th draw of this sample to reduce the autocorrelation arising from the sampling algorithm). Figs. 5 and 6 display the posterior density estimate for the distribution of both the unconstrained and constrained model parameters.

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011

Table 5
Posterior means and standard deviations of the posterior distribution of the MS-GARCH parameters for the S&P500 daily returns.

π_{11}	π_{22}	μ_1	μ_2	γ_1	γ_2	α_1	α_2	β_1	β_2
Daily S&P 500: unconstrained model									
0.986	0.943	0.075	−0.326	0.012	0.183	0.044	0.070	0.927	0.902
(0.003)	(0.011)	(0.016)	(0.058)	(0.003)	(0.040)	(0.008)	(0.020)	(0.010)	(0.027)
Daily S&P 500: constrained model with $\mu_1 = \mu_2 = 0$									
0.987	0.991	–	–	0.026	0.073	0.040	0.092	0.899	0.880
(0.003)	(0.002)	–	–	(0.006)	(0.015)	(0.011)	(0.013)	(0.018)	(0.016)

Notes: Standard errors are in parentheses. The parameter μ_i denotes the conditional mean, γ_i denotes the long-run conditional variance in regime i, α_i and β_i are the regime-specific ARCH and GARCH parameters, respectively, while the π_{ii} are the regime-staying probabilities.

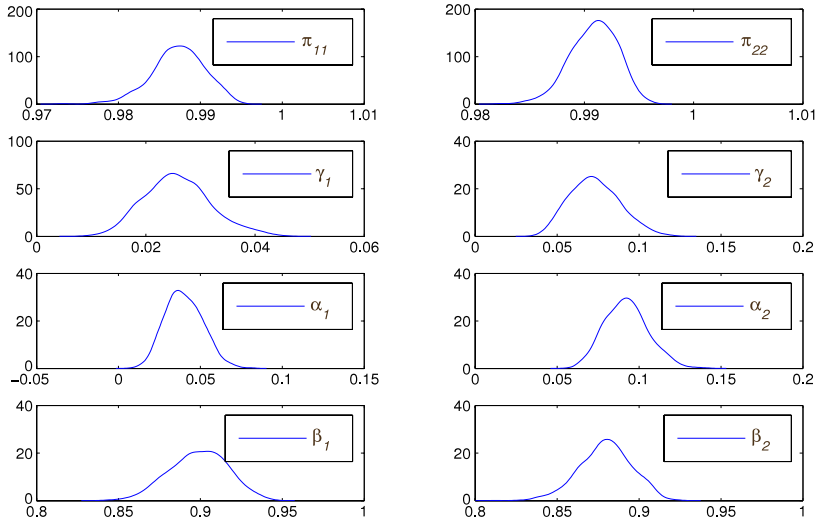


Fig. 5. Posterior densities of the constrained parameters given the data on the S&P 500 daily returns. We apply a multiple-trial metropolized independent sampler (MTMIS) combined with the simplified Klaassen's approximation.

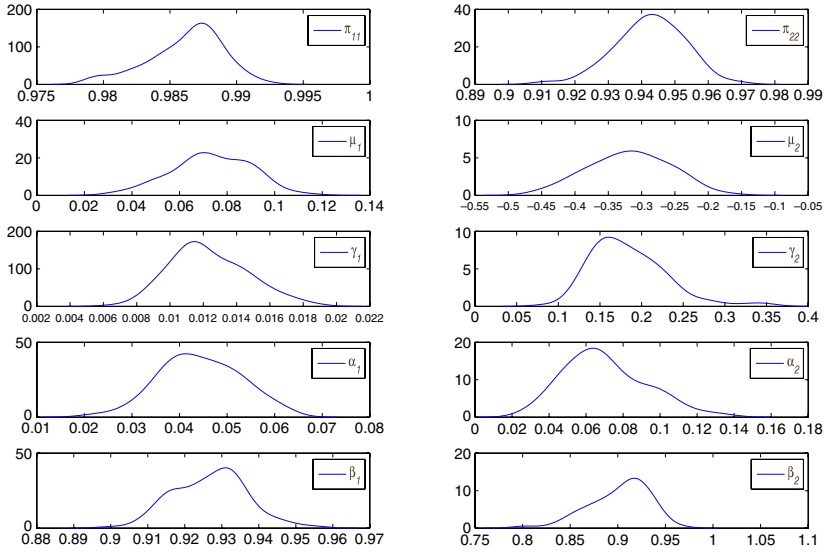


Fig. 6. Posterior densities of the unconstrained parameters given the data on the S&P 500 daily returns. We apply a multiple-trial metropolized independent sampler (MTMIS) combined with the simplified Klaassen's approximation.

The results presented in Table 5 suggest that the estimation of the constrained model is very accurate, but more variability is observed for the unconstrained model. Furthermore, the results of our estimation on the constrained model are consistent with those obtained by Bauwens et al. (2014). In both MS-GARCH specifications, we can identify the first state as a low

volatility regime while the second regime corresponds to a high volatility regime. The second regime of the unconstrained model, however, exhibits (in absolute value) higher mean (μ_2), higher long-run unconditional variance (γ_2), and lower persistence level (i.e., π_{22} is reduced), when compared to the constrained model.

Finally, we assess the performance of our estimation procedure by computing the acceptance rate of the multi-point, multi-move sampling scheme. We obtain a relatively higher acceptance rate of 4% for the unconstrained model as compared to 0.1% for the constrained model. While the acceptance rate of the mean parameters and the GARCH parameters is estimated to be 54% and 25%, respectively. Despite the low acceptance rate for the multi-point proposals, the potentials of the scheme are not undermined as it can be improved upon by increasing the number of trials and the number of the blocks.

6. Conclusion

In this paper, we deal with the challenging issue of designing efficient sampling algorithms for Bayesian inference on Markov-switching GARCH models. We provide some new generalized Metropolis algorithms based on the combination of multi-move and multi-point strategies. Our algorithms extend to Markov-switching nonlinear state space models and the sampling algorithms proposed by So (2006) for continuous nonlinear state space models.

More specifically, we apply the multiple-try sampling strategies of Craiu and Lemieux (2007) with a joint proposal distribution for the hidden states of the Markov-switching GARCH model. For generating candidate paths of the state variable, we apply the Forward Filtering Backward Sampling (FFBS) algorithm to an auxiliary MS-GARCH model. In particular, we propose a unified framework to MS-GARCH approximation that encompasses the existing specifications and provides an avenue to generate new MS-GARCH auxiliary models. Different auxiliary models are considered with respect to the approximation of the GARCH conditional variance equation. We also design a multiple-try algorithm with correlated proposals. To this aim, we introduce an antithetic FFBS based on the permuted displacement method of Craiu and Meng (2005).

We compare our algorithms with simulated data and find that the multiple-trial metropolized independent sampler, in combination with Klaassen's approximation, outperforms the other multi-move, multi-point sampling strategies under consideration. However, due to the cost of coding Klaassen's model, and since there is no clear difference in the parameter estimates obtained by the various sampling algorithms, we suggest the use of the simplified Klaassen's approximation in combination with MTMIS for empirical purposes. Blocking and multiple-trials are also suggested when parameter estimates are close to the boundary.

Acknowledgments

We thank seminar participants at the 5th CSDA International Conference, London, 2013. Our research is supported by funding from the European Union, Seventh Framework Programme FP7/2007–2013 under grant agreement SYRTO-SSH-2012-320270, and by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant MISURA.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.csda.2014.04.011.

References

Abramson, A., Cohen, I., 2007. On the stationarity of Markov-switching GARCH processes. Econom. Theory 23, 485–500.
Ardia, D., 2008. Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications. In: Lecture Notes in Economics and Mathematical Systems, vol. 612. Springer-Verlag, Berlin, Germany.
Arvidsen, N.I., Johnsson, T., 1982. Variance reduction through negative correlation—a simulation study. J. Statist. Comput. Simul. 15, 119–127.
Augustyniak, M., 2013. Maximum likelihood estimation of the Markov-switching GARCH model. Comput. Statist. Data Anal.
Bauwens, L., Dufays, A., Rombouts, J., 2014. Marginal likelihood for Markov-switching and change-point GARCH. J. Econometrics 178, 508–522.
Bauwens, L., Preminger, A., Rombouts, J., 2010. Theory and inference for a Markov switching GARCH model. Econom. J. 13, 218–244.
Besag, J., Green, E., Higdon, D., Mengersen, K., 1995. Bayesian computation and stochastic system (with discussion). Statist. Sci. 10, 3–66.
Bildirici, M., Ersin, O., 2012. Modeling Markov switching ARMA-GARCH neural networks models and an application to forecasting stock returns. Available at SSRN 2125855.
Billio, M., Monfort, A., Robert, C., 1998. The simulated likelihood ratio method. In: Doc Travail CREST. Insee, Paris.
Billio, M., Monfort, A., Robert, C.P., 1999. Bayesian estimation of switching ARMA models. J. Econometrics 93, 229–255.
Bizjajeva, S., Olsson, J., 2008. Antithetic sampling for sequential Monte Carlo methods with application to state space models. Preprints in Mathematical Sciences, Lund University, 14, 1–24.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. J. Econometrics 31, 307–327.
Cai, J., 1994. A Markov model of switching-regime ARCH. J. Bus. Econom. Statist. 12, 309–316.
Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 83, 541–553.
Casarin, R., Craiu, R.V., Leisen, F., 2013. Interacting multiple try algorithms with different proposal distributions. Statist. Comput. 23 (2), 185–200.
Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econometrics 75, 79–97.
Craiu, R.V., Lemieux, C., 2007. Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling. Stat. Comput. 17, 109–120.
Craiu, R.V., Meng, X.L., 2005. Multi-process parallel antithetic coupling for forward and backward MCMC. Ann. Statist. 33, 661–697.
De Jong, P., Shephard, N., 1995. The simulation smoother for time series models. Biometrika 82, 339–350.
Dueker, M., 1997. Markov switching in GARCH processes in mean reverting stock market volatility. J. Bus. Econom. Statist. 15, 26–34.
Dufays, A., 2012. Infinite-state Markov-switching for dynamic volatility and correlation models. CORE discussion paper, 2012/43.

Elliott, R.J., Lau, J.W., Miao, H., Siu, T.K., 2012. Viterbi-based estimation for Markov switching GARCH models. Appl. Math. Finance 19 (3), 1–13.
Fiorentini, G., Planas, C., Rossi, A., 2012. Efficient MCMC sampling in dynamic mixture models. Stat. Comput. 27, 1–13.
Francq, C., Zakoian, J., 2008. Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference. Comput. Statist. Data Anal. 52, 3027–3046.
Frühwirth-Schnatter, S., 1994. Data augmentation and dynamic linear models. J. Time Ser. Anal. 15, 183–202.
Frühwirth-Schnatter, S., 2006. Mixture and Markov-Switching Models. Springer, New York.
Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 4. Oxford University Press, Oxford, pp. 169–193.
Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. J. Financ. Econom. 42, 27–62.
Haas, M., Mittnik, S., Paolella, M., 2004. A new approach to Markov switching GARCH models. J. Financ. Econom. 2, 493–530.
Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. J. Econometrics 64, 307–333.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
He, Z., Maheu, J., 2010. Real time detection of structural breaks in GARCH models. Comput. Statist. Data Anal. 54 (11), 2628–2640.
Henneke, J.S., Rachev, S.T., Fabozzi, F.J., Metodi, N., 2011. MCMC-based estimation of Markov switching ARMA-GARCH models. Appl. Econom. 43 (3), 259–271.
Holmes, C., Jasra, A., 2009. Antithetic methods for Gibbs samplers. J. Comput. Graph. Statist. 18 (2), 401–414.
Jasra, A., Holmes, C., Stephens, D., 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. Statist. Sci. 50–67.
Kaufman, S., Frühwirth-Schnatter, S., 2002. Bayesian analysis of switching ARCH models. J. Time Ser. Anal. 23, 425–458.
Kim, C., Nelson, C., 1999. State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. MIT Press.
Klaassen, F., 2002. Improving GARCH volatility forecasts with regime switching GARCH. Empir. Econom. 27, 363–394.
Koopman, S.J., Durbin, J., 2000. Fast filtering and smoothing for multivariate state space models. J. Time Ser. Anal. 21, 281–296.
Lamoureux, C., Lastrapes, W., 1990. Persistence in variance, structural change, and the GARCH model. J. Bus. Econom. Statist. 8, 225–234.
Liu, J.S., 2002. Monte Carlo Strategies in Scientific Computing. Springer.
Liu, J., Liang, F., Wong, W., 2000. The multiple-try method and local optimization in Metropolis sampling. J. Amer. Statist. Assoc. 95, 121–134.
Liu, J., Wong, W.A.K., 1994. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika 81, 37–40.
Marcucci, J., 2005. Forecasting stock market volatility with regime-switching GARCH models. Stud. Nonlinear Dyn. Econom. 9 (4), 1558–3708.
Marin, J.M., Mengersen, K., Robert, C.P., 2005. Bayesian modelling and inference on mixtures of distributions. Handb. Stat. 25, 459–507.
Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equations of state calculations by fast computing machines. J. Chem. Phys. 1087–1092.
Mikosch, T., Starica, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev. Econom. Stat. 86, 378–390.
Pitt, M.K., Shephard, N., 1996. Antithetic variables for MCMC methods applied to non-Gaussian state space models. In: Proceedings of the Section on Bayesian Statistical Science. Papers presented at the annual meeting of the American Statistical Association, Chicago, IL, USA, August 4–8, 1996 and the International Society for Bayesian Analysis 1996 North American Meeting, Chicago, IL, USA, August 2–3, 1996.
Robert, C., Casella, G., 2010. Monte Carlo Statistical Methods. Springer.
Roberts, G., Sahu, S., 1997. Updating schemes, covariance structure, blocking and parameterisation for the Gibbs sampler. J. R. Stat. Soc. Ser. B 59, 291–318.
So, M., 2006. Bayesian analysis of nonlinear and non-Gaussian state space models via multiple-try sampling methods. Stat. Comput. 16, 125–141.
Wasserman, L., 2000. Asymptotic inference for mixture models using data-dependent priors. J. Amer. Statist. Assoc. 62, 159–180.

Please cite this article in press as: Billio, M., et al., Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.04.011