ARTICLE IN PRESS

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

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Maximum likelihood estimation of the Markov-switching GARCH model

Maciej Augustyniak*

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

ARTICLE INFO

Article history:
Received 16 March 2012
Received in revised form 19 December 2012
Accepted 29 January 2013
Available online xxxx

Keywords:
Markov-switching
GARCH
EM algorithm
Importance sampling

ABSTRACT

The Markov-switching GARCH model offers rich dynamics to model financial data. Estimating this path dependent model is a challenging task because exact computation of the likelihood is infeasible in practice. This difficulty led to estimation procedures either based on a simplification of the model or not dependent on the likelihood. There is no method available to obtain the maximum likelihood estimator without resorting to a modification of the model. A novel approach is developed based on both the Monte Carlo expectation–maximization algorithm and importance sampling to calculate the maximum likelihood estimator and asymptotic variance–covariance matrix of the Markov-switching GARCH model. Practical implementation of the proposed algorithm is discussed and its effectiveness is demonstrated in simulation and empirical studies.

1. Introduction

Financial time series exhibit complex statistical dynamics which are difficult to reproduce with stochastic models. These dynamics are often referred to as the stylized facts of financial data and include, among others, the heavy-tailed nature of the return distribution and volatility clustering (see Cont, 2001). The generalized autoregressive conditional heteroskedasticity (GARCH) class of models (Engle, 1982; Bollerslev, 1986) has been extensively used to model financial data as it offers an explicit way to model volatility. Markov-switching (MS) or regime-switching models have also attracted a lot of attention in the econometric literature since the seminal paper of Hamilton (1989). In MS models the return distribution at a given time depends on the state (or regime) of an unobserved Markov chain. The states of the Markov chain are often given an economic interpretation. For example, a regime with a negative mean return and high volatility may be associated with a state of financial distress in the economy.

Due to the popularity of MS and GARCH models, it is natural to combine these two approaches and consider a MS-GARCH model. The MS-GARCH model can be simply understood as a GARCH model where parameters depend on the state of an unobserved Markov chain. One way to justify such a combination is given by Lamoureux and Lastrapes (1990) and Mikosch and Starica (2004) who show that the high persistence observed in the variance of financial returns can be explained by time-varying GARCH parameters.

Hamilton and Susmel (1994) were among the first authors to discuss the MS-GARCH model. They noted that the estimation of this path dependent model is a challenging task because exact computation of the likelihood is infeasible in practice. This led some authors (Dueker, 1997; Gray, 1996; Haas et al., 2004; Klaassen, 2002) to propose estimating modified versions of the MS-GARCH model that circumvent the path dependence problem by maximum likelihood. Other authors suggested alternative estimation methods such as a generalized method of moments (GMM) procedure (Francq and

* Tel.: +1 514 5654726.
E-mail address: augusty@dms.umontreal.ca.
URL: http://www.dms.umontreal.ca/~augusty/.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.01.026

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Zakoïan, 2008) and a Bayesian Markov chain Monte Carlo (MCMC) algorithm (Bauwens et al., 2010, 2011). To this date, there is no method available to obtain the maximum likelihood estimator (MLE) of the MS-GARCH model without resorting to a simplification of the model.

The objective and main contribution of this article is to develop a novel approach based on the Monte Carlo expectation–maximization (MCEM) algorithm (Wei and Tanner, 1990) and the Monte Carlo maximum likelihood (MCML) method (Geyer, 1994, 1996) to estimate the MLE of the MS-GARCH model. The proposed algorithm requires simulations from the posterior distribution of the state vector. For this reason, it can be seen as a frequentist counterpart of the Bayesian MCMC method proposed by Bauwens et al. (2010) in the sense that both algorithms build on the data augmentation technique (Tanner and Wong, 1987). A secondary contribution of this article is to show how the asymptotic variance–covariance matrix of the MLE can be estimated. This is relevant since Francq and Zakoïan (2008) were not able to obtain the asymptotic standard errors of their GMM estimates due to numerical difficulties.

This paper is organized as follows. Section 2 defines the MS-GARCH model. Section 3 introduces the novel approach to calculate the MLE, proposes a procedure to approximate the asymptotic variance–covariance matrix of the MLE and discusses practical implementation of the algorithm. Section 4 demonstrates the effectiveness of the proposed method in a simulation study. Section 5 applies the estimation technique to daily and weekly log-returns on the S&P 500 index. Section 6 concludes and proposes avenues for further research. Moreover, Appendix A justifies the validity of the expectation–maximization (EM) algorithm when applied to the MS-GARCH model. Appendices B and C include a proof and some technical details related to the implementation of the algorithm.

2. The MS-GARCH model

2.1. Definition

Following Bauwens et al. (2010) and Francq et al. (2001), the MS-GARCH model can be defined by the following equations:

$$y_t = \mu_{S_t} + \sigma_t(S_{1:t})\eta_t, \tag{1}$$

$$\sigma_t^2(S_{1:t}) = \omega_{S_t} + \alpha_{S_t}\epsilon_{t-1}^2(S_{t-1}) + \beta_{S_t}\sigma_{t-1}^2(S_{1:t-1}), \tag{2}$$

$$\epsilon_{t-1}(S_{t-1}) = y_{t-1} - \mu_{S_{t-1}}. \tag{3}$$

The vector $(y_1, \ldots, y_T)$ represents the observations to be modeled and η_t, $t = 1, \ldots, T$, are independent and identically distributed normal innovations with zero mean and unit variance. At each time point, the conditional mean of the observation y_t is $\mu_{S_t} = \mathrm{E}(y_t \mid S_t)$ and the conditional variance is $\sigma_t^2(S_{1:t}) = \mathrm{Var}(y_t \mid y_{1:t-1}, S_{1:t})$, where $y_{1:t-1}$ and $S_{1:t}$ are shorthand for the vectors $(y_1, \ldots, y_{t-1})$ and $(S_1, \ldots, S_t)$, respectively. The process $\{S_t\}$ is an unobserved ergodic time-homogeneous Markov chain with N-dimensional discrete state space (i.e., S_t can take integer values from 1 to N). The $N \times N$ transition matrix of the Markov chain is defined by the transition probabilities $\{p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)\}_{i,j=1}^N$. The vector $\theta = (\{\mu_i, \omega_i, \alpha_i, \beta_i\}_{i=1}^N, \{p_{ij}\}_{i,j=1}^N)$ denotes the parameters of the model. To ensure positivity of the variance, the following constraints are required: $\omega_i > 0, \alpha_i \geq 0$ and $\beta_i \geq 0, i = 1, \ldots, N$. Since $\sum_{j=1}^N p_{ij} = 1$ for $i = 1, \ldots, N$, θ contains $(4N + N(N-1))$ free parameters. Conditions for stationarity and the existence of moments were studied by Bauwens et al. (2010), Francq et al. (2001) and Francq and Zakoïan (2005).

2.2. Path dependence problem

The specification (1)–(3) causes difficulties in estimation since the conditional variance at time t depends on the entire regime path $S_{1:t}$. To emphasize this dependence, the notation $\sigma_t^2(S_{1:t})$ is used in Eqs. (1)–(3), but to simplify it in what follows, σ_t^2 will be used to represent $\sigma_t^2(S_{1:t})$. Moreover, let y and S denote $y_{1:T}$ and $S_{1:T}$, respectively, and $f(p)$ stand for a probability density (mass) function. The calculation of the likelihood of the observations, denoted by $f(y \mid \theta)$, can be accomplished by integrating out all possible regime paths:

$$\begin{aligned} f(y \mid \theta) &= \sum_S f(y, S \mid \theta) = \sum_S f(y \mid S, \theta) p(S \mid \theta) \\ &= \sum_S \left[\prod_{t=1}^T \sigma_t^{-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_t - \mu_{S_t})^2}{2\sigma_t^2} \right) \right] p(S \mid \theta). \end{aligned} \tag{4}$$

For large T, this integration is infeasible numerically as the summation in Eq. (4) contains N^T terms and quickly becomes very large. Even the estimation of the likelihood by brute force Monte Carlo (i.e., by simulating independent sequences of states from the underlying Markov chain) will fail since such estimators exhibit prohibitively large variances (see Danielsson and Richard, 1993). Nevertheless, as shown by Bauwens et al. (2011), it is possible to obtain an accurate estimate of the log-likelihood by writing

$$\log f(y \mid \theta) = \log f(y_1 \mid \theta) + \sum_{t=1}^{T-1} \log f(y_{t+1} \mid y_{1:t}, \theta),$$

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

and estimating $f(y_{t+1} \mid y_{1:t}, \theta)$, $t = 1, \ldots, T-1$, sequentially with the aid of particle filters. Unfortunately, the estimate of the log-likelihood obtained with particle filters is not a continuous function of θ (see Pitt, 2002). Therefore, this simulated log-likelihood is difficult to maximize with standard optimization routines.

2.3. The solution of Gray (1996)

Gray (1996) was the first to suggest a method to estimate the MS-GARCH model. Recognizing that the likelihood cannot be computed exactly in practice, he proposed to replace Eqs. (2) and (3) with:

$$\sigma_t^2 = \omega_{S_t} + \alpha_{S_t}\epsilon_{t-1}^2 + \beta_{S_t} h_{t-1},$$
$$\epsilon_{t-1} = y_{t-1} - \mathrm{E}(y_{t-1} \mid y_{1:t-2}),$$

where $h_{t-1} = \mathrm{Var}(y_{t-1} \mid y_{1:t-2})$. The expression h_{t-1} has the effect of collapsing all of the possible conditional variances at time $t-1$ into a single value that does not depend on the regime path. As a consequence, the conditional distribution of y_t, $f(y_t \mid y_{1:t-1}, S_{1:t}, \theta)$, is now independent of $S_{1:t-1}$ and maximum likelihood estimation is tractable (see Hamilton, 2008). Dueker (1997) and Klaassen (2002) expanded on Gray's idea by using broader information sets to collapse variances. The quality of these approximations to estimate the MLE of the path dependent MS-GARCH model has not been investigated. Intuitively, Gray's method will be more reliable if regimes can be inferred accurately from the data. For example, this occurs when regimes are persistent (i.e., p_{ii}, $i = 1, \ldots, N$, are close to 1) and well differentiated in their parameters. However, even in this case, the simulation study in Section 4 suggests that Gray's method does not generate consistent estimates for the path dependent MS-GARCH model. Therefore, an alternative approach to compute the MLE is needed.

3. A novel approach to estimate the MS-GARCH model

3.1. MCEM algorithm

There are many situations in statistical inference where it is difficult to maximize the likelihood of the observed data directly. The EM algorithm is a technique designed to obtain the MLE of the observed data likelihood through an iterative procedure that does not require the computation of the likelihood. Instead, we must be able to calculate and maximize

$$\begin{aligned} Q(\theta \mid \theta') &= \mathrm{E}\left(\log[f(y, S \mid \theta)] \mid y, \theta'\right) \\ &= \sum_S \log[f(y, S \mid \theta)] p(S \mid y, \theta'). \end{aligned} \tag{5}$$

The complete data likelihood of the MS-GARCH model, $f(y, S \mid \theta)$, admits a simple expression (see Eq. (4)) but it is not possible to calculate (5) exactly because of the path dependence problem. Nonetheless, we can simulate from $p(S \mid y, \theta')$ using Gibbs sampling and obtain a Monte Carlo approximation of $Q(\theta \mid \theta')$. When the expectation step of the EM algorithm is approximated, we obtain the MCEM algorithm introduced by Wei and Tanner (1990). In such cases, the monotonicity property of the EM algorithm (i.e., the likelihood of the observed data is never decreased at each iteration) is not guaranteed to hold because of the Monte Carlo error introduced. Consequently, the specification of the number of simulated state vectors at each iteration of the algorithm is of central importance. Wei and Tanner (1990) recommend that small values be used in the initial stages and that these values be increased as the algorithm moves closer to convergence. The validity of the EM algorithm in the context of the MS-GARCH model is discussed in Appendix A.

3.2. MCML algorithm

A common criticism of the EM algorithm is that, although it can reach the neighborhood of the MLE quickly, it exhibits slow linear convergence in the neighborhood itself (see McLachlan and Krishnan, 2008, Section 3.9). For this reason, it is often suggested to combine the EM algorithm with a Newton–Raphson method or to simply switch to a faster method after a few EM iterations. The latter was suggested by McCulloch (1997) who proposed to follow the MCEM algorithm with the MCML approach of Geyer (1994, 1996). Suppose that $\{S^{(i)}\}_{i=1}^{m^\star}$ are simulated state vectors from $p(S \mid y, \theta^\star)$ and define

$$w_{\theta|\theta^\star}^{(i)} = \frac{f(y, S^{(i)} \mid \theta)}{f(y, S^{(i)} \mid \theta^\star)}, \quad i = 1, \ldots, m^\star. \tag{6}$$

The MCML algorithm makes use of importance sampling to directly maximize the log-likelihood through the following relation:

$$\log f(y \mid \theta) - \log f(y \mid \theta^\star) = \log \mathrm{E}\left[\frac{f(y, S \mid \theta)}{f(y, S \mid \theta^\star)} \mid y, \theta^\star\right] \approx \log \frac{1}{m^\star} \sum_{i=1}^{m^\star} w_{\theta|\theta^\star}^{(i)}. \tag{7}$$

Due to the well-known problems related to importance sampling, Cappé et al. (2002) explain that this method does not work well unless $\theta^\star$ is in a close neighborhood of the MLE. Similarly to McCulloch (1997), they suggest to use it only with another consistent maximum likelihood estimation method.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

3.3. MCEM–MCML algorithm for the MS-GARCH model

The discussion in Section 3.2 suggests a hybrid MCEM–MCML algorithm. First, iterations of the MCEM algorithm can be performed to get a good estimate, $\theta^\star$, of the MLE. This estimate is then used to generate the importance sample in the MCML algorithm. Both algorithms complement each other: the MCEM algorithm addresses the flaw of the MCML algorithm relating to the choice of $\theta^\star$ while the MCML method replaces many potential MCEM iterations with a single iteration, leading to a faster convergence.

Given an initial guess $\theta^{(0)}$, the following algorithm started at $r = 1$ produces a sequence of iterates $\{\theta^{(r)}\}_{r\geq 1}$ allowing us to compute the MLE of model (1)–(3):

Algorithm 1 (*MCEM–MCML Algorithm*)**.**

1. Simulate m_r samples of the state vector S from $p(S \mid y, \theta^{(r-1)})$ using a single-move Gibbs sampler. The states are simulated sequentially for $t = 1, \ldots, T$ based on the following full conditional distribution:

$$p(S_t \mid S^{(i)}_{1:t-1}, S^{(i-1)}_{t+1:T}, y, \theta^{(r-1)}) \propto p_{S^{(i)}_{t-1}, S_t}\, p_{S_t, S^{(i-1)}_{t+1}} \prod_{j=t}^{T} \sigma_j^{-1} \exp\left(-\frac{(y_j - \mu_{S_j})^2}{2\sigma_j^2}\right). \tag{8}$$

   To ease notation, the expression $\sigma_j(S_{1:t})$ was reduced to σ_j. In the context of Eq. (8), σ_j represents $\sigma_j(S^{(i)}_{1:t-1}, S_t, S^{(i-1)}_{t+1:j})$. It is straightforward to sample S_t from (8) since S_t can only take integer values from 1 to N. However, it should be noted that it is not possible to compute expression (8) numerically for each value of S_t since this will result in underflow. To avoid underflow, we can calculate the ratios of these expressions and then recover the probabilities for $S_t = 1, \ldots, N$ from them. The m_r simulations of the state vector S that are obtained are denoted by $\{S^{(i)}\}_{i=1}^{m_r}$. These draws form a Markov chain with $p(S \mid y, \theta^{(r-1)})$ as its stationary distribution (see Frühwirth-Schnatter, 2006, Section 3.4.1).
2. Monte Carlo E-step: Calculate $\widehat{Q}(\theta \mid \theta^{(r-1)})$, an approximation of the conventional E-step $Q(\theta \mid \theta^{(r-1)})$, where

$$\widehat{Q}(\theta \mid \theta^{(r-1)}) = \frac{1}{m_r}\sum_{i=1}^{m_r} \log[f(y, S^{(i)} \mid \theta)] \tag{9}$$

$$= -\frac{T\log(2\pi)}{2} - \frac{1}{2m_r}\sum_{t=1}^{T}\sum_{i=1}^{m_r}\left[\log(\sigma_t^{(i)})^2 + \frac{(y_t - \mu_{S_t^{(i)}})^2}{(\sigma_t^{(i)})^2}\right] + \frac{1}{m_r}\sum_{t=1}^{T}\sum_{i=1}^{m_r}\log(p_{S^{(i)}_{t-1}, S^{(i)}_t}) \tag{10}$$

$$= \text{term 1} + \text{term 2}.$$

   In the previous expressions, $\sigma_t^{(i)}$ is shorthand for $\sigma_t(S^{(i)}_{1:t})$.
3. M-step: Perform the following maximization:

$$\theta^{(r)} = \arg\max_{\theta} \widehat{Q}(\theta \mid \theta^{(r-1)}).$$

   This optimization can be split into two independent steps since terms 1 and 2 of Eq. (10) involve different subsets of the parameters. Term 1 includes the mean and GARCH parameters while term 2 only contains transition probabilities. Maximization of term 1 must be performed numerically and is similar to a standard GARCH optimization to calculate the MLE. To improve the performance of that optimization, the gradient of term 1 with respect to the mean and GARCH parameters should be provided to the optimization routine (see Appendix C). Maximization of term 2 can be done analytically. Term 2 is at its maximum when the transition probabilities take the values

$$p_{jk} = \frac{f_{jk}}{\sum_{l=1}^{N} f_{jl}}, \quad j, k = 1, \ldots, N,$$

   where f_{jk} denotes the total number of transitions from state j to state k in all of the m_r simulated state vectors. A proof of this result is in Appendix B.
4. Apply a decision rule to determine whether to switch to the MCML algorithm (see Section 3.5.2). If the decision is to switch, go to step 5 and set $\theta^\star = \theta^{(r)}$. Otherwise, add 1 to r and go to step 1.
5. Simulate $m^\star$ samples of the state vector S from $p(S \mid y, \theta^\star)$ using the single-move Gibbs sampler described in step 1 of the algorithm to obtain the importance sample $\{S^{(i)}\}_{i=1}^{m^\star}$.
6. MCML-step: Perform the following maximization to obtain the MLE:

$$\hat{\theta} = \arg\max_{\theta}\left[\log\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\right]. \tag{11}$$

   In contrast to the M-step, this optimization cannot be split into two steps. Appendix C provides some details related to its implementation.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Using importance sampling, the final sample, $\{S^{(i)}\}_{i=1}^{m^\star}$, generated at step 5 of the algorithm can be transformed into a weighted sample, $\{S^{(i)}, \bar{w}^{(i)}_{\hat{\theta}|\theta^\star}\}_{i=1}^{m^\star}$, from $p(S \mid y, \hat{\theta})$, where $\bar{w}^{(i)}_{\hat{\theta}|\theta^\star} = w^{(i)}_{\hat{\theta}|\theta^\star} / \sum_{i=1}^{m^\star} w^{(i)}_{\hat{\theta}|\theta^\star}$, $i = 1, \ldots, m^\star$. This sample can be used to obtain an estimate of the smoothed inference of the state at time t, $p(S_t = j \mid y, \hat{\theta})$, $j = 1, \ldots, N$, with $\sum_{i=1}^{m^\star} \bar{w}^{(i)}_{\hat{\theta}|\theta^\star} I_{\{S_t^{(i)}=j\}}$ or to compute the asymptotic variance–covariance matrix of the MLE.

3.4. Asymptotic variance–covariance matrix of the MLE

The asymptotic variance–covariance matrix of the MLE for the MS-GARCH model can be obtained from the inverse of the Fisher information matrix, denoted by $\mathcal{F}(\theta) = \mathrm{E}\left[\mathcal{s}(\theta)\mathcal{s}(\theta)^T\right]$, where $\mathcal{s}(\theta) = \partial \log f(y \mid \theta)/\partial\theta$ is the score related to the observed data log-likelihood. The expectation is taken over y and $\mathcal{F}(\theta)$ must be evaluated at the true parameter values (not the MLE). We may approximate $\mathcal{F}(\theta)$ by generating independent realizations of y and averaging $\mathcal{s}(\theta)\mathcal{s}(\theta)^T$ over all of these realizations. Unfortunately, it is not possible to calculate the score exactly, but we may approximate it using the following relation (see McLachlan and Krishnan, 2008, Section 3.7):

$$\mathcal{s}(\theta) = \left[\frac{\partial}{\partial\theta'} Q(\theta' \mid \theta)\right]_{\theta'=\theta} \approx \left[\frac{\partial}{\partial\theta'} \widehat{Q}(\theta' \mid \theta)\right]_{\theta'=\theta}, \tag{12}$$

where $\partial\widehat{Q}(\theta' \mid \theta)/\partial\theta'$ is the gradient of the Monte Carlo E-step which is available in closed form (see Appendix C). An alternative to approximate the score is to use the MCML relation in expression (7). If $\theta^\star = \theta$, this alternative is equivalent to using expression (12). The procedure just described generates a valid approximation of the asymptotic variance–covariance matrix of the MLE when the true parameters of the MS-GARCH model are known. This is the method that is used to obtain the asymptotic standard errors of the MLE in the simulation study of Section 4.

When a real data set is fitted to the MS-GARCH model, the true parameters are not known and we have two choices to approximate the variance–covariance matrix of the MLE. First, we may replace the true parameters with the MLE in the method just mentioned. However, it is common practice to estimate the variance–covariance matrix with the inverse of $J(\hat{\theta})$, the observed information matrix evaluated at the MLE. For the MS-GARCH model, $J(\hat{\theta})$ can be approximated based on the decomposition of the observed information matrix presented by Louis (1982) (see also McLachlan and Krishnan, 2008, Sections 4.2.2 and 6.3.5):

$$\begin{aligned} J(\hat{\theta}) &= \left[-\frac{\partial^2}{\partial\theta\partial\theta^T} \log f(y \mid \theta)\right]_{\theta=\hat{\theta}} \\ &= \left[-\frac{\partial^2}{\partial\theta\partial\theta^T} Q(\theta \mid \hat{\theta})\right]_{\theta=\hat{\theta}} - \mathrm{E}\left[\mathcal{s}_c(\hat{\theta})\mathcal{s}_c(\hat{\theta})^T \mid y, \hat{\theta}\right] + \mathcal{s}(\hat{\theta})\mathcal{s}(\hat{\theta})^T, \end{aligned} \tag{13}$$

where $\mathcal{s}_c(\hat{\theta}) = [\partial \log f(y, S \mid \theta)/\partial\theta]_{\theta=\hat{\theta}}$ is the score related to the complete data log-likelihood evaluated at the MLE. Since $\mathcal{s}(\hat{\theta})$ will be very close to the zero vector in practice, it is only the first two terms in expression (13) that need to be estimated. This approximation can be performed using the final state vectors generated at step 5 of the MCEM–MCML algorithm. As mentioned at the end of Section 3.3, $\{S^{(i)}, \bar{w}^{(i)}_{\hat{\theta}|\theta^\star}\}_{i=1}^{m^\star}$ is a weighted sample from $p(S \mid y, \hat{\theta})$. Hence, we can write:

$$\begin{aligned} \frac{\partial^2}{\partial\theta\partial\theta^T} Q(\theta \mid \hat{\theta}) &= \sum_S \left[\frac{\partial^2}{\partial\theta\partial\theta^T} \log[f(y, S \mid \theta)]\right] p(S \mid y, \hat{\theta}) \\ &\approx \sum_{i=1}^{m^\star} \left[\frac{\partial^2}{\partial\theta\partial\theta^T} \log[f(y, S^{(i)} \mid \theta)]\right] \bar{w}^{(i)}_{\hat{\theta}|\theta^\star}. \end{aligned} \tag{14}$$

Expression (14) can be computed by numerical differentiation of

$$\sum_{i=1}^{m^\star} \left[\frac{\partial}{\partial\theta} \log[f(y, S^{(i)} \mid \theta)]\right] \bar{w}^{(i)}_{\hat{\theta}|\theta^\star},$$

which is available in closed form (see Appendix C). Moreover, the second term in Eq. (13) can be estimated with

$$\mathrm{E}\left[\mathcal{s}_c(\theta)\mathcal{s}_c(\theta)^T \mid y, \hat{\theta}\right] \approx \sum_{i=1}^{m^\star} \left[\frac{\partial}{\partial\theta} \log[f(y, S^{(i)} \mid \theta)]\right] \left[\frac{\partial}{\partial\theta} \log[f(y, S^{(i)} \mid \theta)]\right]^T \bar{w}^{(i)}_{\hat{\theta}|\theta^\star}.$$

Therefore, the estimation of the variance–covariance matrix of the MLE with $[J(\hat{\theta})]^{-1}$ can be obtained as a by-product of the MCEM–MCML algorithm.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Table 1
Simulation schedule 1.

m_1	m_2	m_3	m_4	$m^\star$
500	1000	2500	5000	10 000

Table 2
Simulation schedule 2.

m_1 to m_{10}	m_{11} to m_{28}	m_{29}	m_{30}	$m^\star$
500	1000	2500	5000	40 000

3.5. Practical considerations with regard to the MCEM–MCML algorithm

Section 3.3 shows how the MCEM and MCML algorithms can be used to estimate the MLE of the MS-GARCH model but it does not convey any guidance on how to choose the number of simulated state vectors at each iteration, nor does it tell us when to switch to the MCML algorithm. Starting values for the algorithm and the Gibbs sampler are also of concern.

3.5.1. Starting values

For the MCEM algorithm and the Gibbs sampler, starting values are important since they can influence convergence. Section 4 demonstrates that the MLE of Gray's model (see Section 2.3) provides a good initial value for the MCEM–MCML algorithm. Nevertheless, it is always prudent to try a few additional starting points to avoid suboptimal convergence (see Appendix A). For example, the transition probabilities and mean parameters can be set equal to those estimated for a basic MS model while the volatility parameters can be derived from a fitted standard GARCH model. The approximations of Dueker (1997) or Klaassen (2002) can also be considered, but they will generally yield initial values in the same range as Gray's model.

To initialize the Gibbs sampler at the first iteration of the MCEM algorithm we require a vector of states. This can be obtained from smoothed inferences of the states using Gray's model which can be calculated recursively (see Hamilton, 1994, p. 694). For instance, for each $t = 1, \ldots, T$, we can compute the smoothed inference $p(S_t \mid y, \theta^{(0)})$ and pick the value of S_t which is most likely. This collection of states can form the initial state vector. Moreover, from one iteration of the MCEM algorithm to the next we may reuse the last state vector generated on a given iteration as the starting state vector on the next iteration. Since a reasonable starting value for the state vector is available, it is not strictly necessary to use a burn-in sample (see Jones and Hobert, 2001).

A final technical detail is the choice of S_0. To generate the first state of the Markov chain S_1, we need to make an assumption about S_0. We have at least two choices. First, we may assume that the Markov chain $\{S_t\}$ is stationary and therefore S_1 can be generated according to the stationary distribution of the Markov chain. Second, we may treat the initial distribution of S_0, say the vector δ, as a parameter of the model and estimate it. However, maximizing the likelihood over δ is equivalent to maximizing the likelihood conditionally on starting from the N possible different states. In other words, the MLE of δ is a unit vector of the form $(0, \ldots, 0, 1, 0, \ldots, 0)$ (see Zucchini and MacDonald, 2009, Section 4.2.4). For simplicity and since this assumption does not play a material role in estimation, I will suppose that the initial state S_0 is given and fixed instead of needing estimation.

3.5.2. Simulation schedule

A strategy for increasing the sample size throughout the MCEM–MCML algorithm is proposed in this section and referred to as the simulation schedule of the algorithm. To determine an appropriate simulation schedule, I applied the MCEM–MCML algorithm with an excessive amount of simulations and iterations on the MS-GARCH model considered in Section 4 and on the empirical data studied in Section 5. Afterwards, I reduced the number of simulations and compared the ending parameter vectors to the ones obtained by brute force.

I recommend two simulation schedules for the MCEM–MCML algorithm (see Tables 1 and 2). Simulation schedule 1 permits a fast estimation (see Section 4.4 for computational times) while simulation schedule 2 puts more emphasis on accuracy and is more robust with respect to the choice of starting values. In the simulation study (Section 4), simulation schedule 1 is used since simulation schedule 2 does not offer significant improvements in the estimation process. However, when considering empirical data in Section 5, Simulation schedule 2 is preferred because it yields important gains in precision. For example, different runs of the MCEM–MCML algorithm conducted with this strategy brought the parameter iterates to a close neighborhood of the ones obtained by brute force (see Section 5.1).

The effectiveness of a given simulation schedule depends on the data generating process, the sample generated and starting values. Consequently, it may be worthwhile to automate the MCEM–MCML algorithm, i.e., dictate rules that automatically select the sample size at each iteration and determine when to perform the final MCML iteration. For example, we may want to switch to the MCML algorithm if the relative differences between successive MCEM parameter iterates are below a certain threshold. There are at least two good papers dealing with the automation of the MCEM algorithm when

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

MCMC samples are used. On one hand, Levine and Fan (2004) propose to select the sample size at each iteration based on a confidence ellipsoid created around the parameter iterates. On the other hand, Caffo et al. (2005) base that decision on whether the Monte Carlo E-step of the algorithm actually increased the likelihood with high probability. Both of these approaches rely on asymptotic results and require subsampling approximately independent subsets of the MCMC sample. As a consequence, it is not guaranteed that they can be used with high reliability. Additionally, a certain amount of manual adjustments will always be necessary. For these reasons, automated strategies were not implemented. Nevertheless, it would be interesting to see how these strategies compare to the ones that were recommended. This element is open for future research.

4. Simulation study

This section evaluates the effectiveness of the proposed MCEM–MCML algorithm to compute the MLE of the MS-GARCH model. The algorithm was coded to work with version 2.15.0 of the R software (R Development Core Team, 2012). The Gibbs sampler involves nested sequential loops which render the generation of states a slow process with R. Consequently, I used the Rcpp package (Eddelbuettel and François, 2011) available with R to implement the generation of states in C++ (the Rcpp package simplifies the integration of C++ code with R). All results presented in this section are reproducible since I seeded the random number generator in R. The code to reproduce these results is available on my website.

4.1. Description

I simulated 200 independent trajectories of the MS-GARCH model with sizes of $T = 500$, 1500 and 5000 based on the parameter set presented in the third column of Table 3. All of the simulations were started assuming an initial state value of $S_0 = 1$ and an initial variance of $\sigma_0^2 = 2.56$ (this value is approximately equal to the unconditional variance of the process). This parameter set (denoted by BPR) was considered by Bauwens et al. (2010) to assess their Bayesian MCMC algorithm. Bauwens et al. (2010) state that their choice of parameters is inspired by empirical studies. It is thus a reasonable representation of a process for generating financial returns. In particular, the first regime defines a positive mean return–low volatility environment while the second regime pertains to a negative mean return–high volatility state.

For each of the simulated trajectories of the MS-GARCH model, the MLE was estimated by applying the MCEM–MCML algorithm with simulation schedule 1 (see Table 1). To examine the effect of starting values, the iterations were started from the true parameter values and from the MLE of the model of Gray (1996). It is not possible to compare the MCEM–MCML algorithm to the GMM algorithm of Francq and Zakoïan (2008) on parameter set BPR because the GMM algorithm was developed on a restricted version of model (1)–(3) with zero means in all regimes. It seems difficult to extend this algorithm to the case of regime-dependent means or to more general versions of the MS-GARCH model.

4.2. Identifiability issues

MS models are generally not identifiable since the parameters of two states may be permuted without changing the likelihood. When Bayesian MCMC techniques are used to estimate MS models, the label-switching problem (see Frühwirth-Schnatter, 2006, Section 3.5.5) can seriously complicate statistical inference. This problem occurs because the posterior distribution of the parameters is invariant under a permutation of state indices when exchangeable priors are used. Therefore, the MCMC output is difficult to interpret when one needs to draw inferences about parameters in specific regimes. A similar problem occurs in a frequentist setting when we want to evaluate the sampling distribution of estimators in a simulation study. We must associate each estimated parameter to a state index since the objective is to assess biases and standard errors of the estimated parameters in a given regime. In contrast to the label-switching problem, this problem received very little attention (see Yao, submitted for publication). The approach followed here to identify regimes is to impose an identifiability constraint of the form $\omega_1 < \omega_2 < \cdots < \omega_N$. This is a simple solution that works well here since the parameters ω_1 and ω_2 differ significantly between regimes in parameter set BPR. An alternative would be to impose a similar constraint on the means, but there is less of a clear-cut difference between the two regimes based on that choice. If the regimes are not well separated, such an approach may not correctly identify the estimated parameters with their corresponding regimes. More sophisticated approaches have been proposed to deal with the label-switching problem, but they are not always applicable in a frequentist setting.

It must be emphasized that when a data set is fitted with the MCEM–MCML algorithm, the label-switching problem is not encountered since the goal is to find a point estimate corresponding to one of the equivalent modes of the likelihood function. It also does not cause a problem for the estimation of the asymptotic variance–covariance matrix. This problem is only encountered when evaluating the sampling distribution of estimators.

4.3. Discussion of results

Table 3 displays the summary statistics for 200 estimates of the MLE obtained with Gray's model (denoted by Gray), the MCEM–MCML algorithm started from the MLE of Gray's model (denoted by MCEM_G) and the MCEM–MCML algorithm

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Table 3
Mean and RMSE based on 200 estimates of the MLE.

T		Value	A-StErr	Mean			RMSE		
				Gray	MCEM_G	MCEM_0	Gray	MCEM_G	MCEM_0
500	μ_1	0.06	0.039	0.061	0.059	0.060	0.059	0.075	0.070
	μ_2	−0.09	0.211	−0.080	−0.091	−0.093	0.216	0.204	0.212
	ω_1	0.30	0.096	0.329	0.298	0.300	0.092	0.091	0.091
	α_1	0.35	0.110	0.356	0.336	0.342	0.146	0.150	0.146
	β_1	0.20	0.162	0.106	0.211	0.201	0.184	0.197	0.182
	ω_2	2.00	1.539	3.684	2.786	2.630	2.324	1.852	1.720
	α_2	0.10	0.106	0.118	0.121	0.113	0.152	0.152	0.135
	β_2	0.60	0.279	0.354	0.425	0.469	0.391	0.365	0.339
	p_{11}	0.98	0.010	0.976	0.976	0.977	0.017	0.017	0.013
	p_{22}	0.96	0.023	0.952	0.954	0.954	0.028	0.029	0.026
1500	μ_1	0.06	0.025	0.062	0.061	0.061	0.024	0.023	0.023
	μ_2	−0.09	0.114	−0.087	−0.091	−0.093	0.111	0.114	0.115
	ω_1	0.30	0.054	0.345	0.300	0.301	0.069	0.058	0.059
	α_1	0.35	0.071	0.362	0.350	0.351	0.069	0.066	0.065
	β_1	0.20	0.091	0.081	0.201	0.199	0.131	0.103	0.103
	ω_2	2.00	1.006	3.733	2.304	2.325	1.987	1.193	1.171
	α_2	0.10	0.060	0.101	0.088	0.089	0.061	0.057	0.058
	β_2	0.60	0.187	0.376	0.562	0.559	0.269	0.198	0.192
	p_{11}	0.98	0.006	0.979	0.979	0.979	0.007	0.007	0.007
	p_{22}	0.96	0.011	0.957	0.958	0.958	0.015	0.014	0.015
5000	μ_1	0.06	0.012	0.062	0.061	0.061	0.014	0.013	0.013
	μ_2	−0.09	0.065	−0.081	−0.090	−0.090	0.061	0.061	0.061
	ω_1	0.30	0.032	0.351	0.300	0.301	0.057	0.028	0.027
	α_1	0.35	0.037	0.362	0.354	0.352	0.041	0.038	0.038
	β_1	0.20	0.052	0.072	0.198	0.196	0.131	0.050	0.049
	ω_2	2.00	0.563	3.638	2.051	2.060	1.721	0.554	0.532
	α_2	0.10	0.030	0.112	0.093	0.094	0.038	0.035	0.034
	β_2	0.60	0.101	0.385	0.599	0.597	0.234	0.103	0.098
	p_{11}	0.98	0.003	0.980	0.980	0.980	0.003	0.003	0.003
	p_{22}	0.96	0.006	0.958	0.959	0.959	0.007	0.007	0.006

Table 4
Mean values of 1000 estimates of the MLE for a GARCH model.

T	μ	ω	α	β
500	−0.090	2.357	0.104	0.541
1500	−0.091	2.220	0.102	0.565
5000	−0.091	2.051	0.100	0.592
10 000	−0.089	2.024	0.100	0.596
50 000	−0.090	1.999	0.100	0.600
True value	−0.090	2.000	0.100	0.600

started from the true parameter values (denoted by MCEM_0) for each of the sample sizes considered ($T = 500$, 1500 and 5000). RMSE denotes the root-mean-square error and A-StErr stands for asymptotic standard error.

We can observe that the estimates of ω_1, ω_2, β_1 and β_2 based on Gray's model display significant biases which do not necessarily decrease as the sample size is increased. This suggests that Gray's model does not generate consistent estimates for the path dependent MS-GARCH model. However, it does provide a good starting value for the MCEM–MCML algorithm as results under columns MCEM_G and MCEM_0 are very close to each other. In fact, the MCEM–MCML algorithm started from the MLE of Gray's model or from the true parameter values generally converged to the same mode of the likelihood function.

The results shown in Table 3 corroborate that the MCEM–MCML algorithm started from the MLE of Gray's model is an effective method to estimate the MS-GARCH model. The following remarks support this assertion.

1. The RMSE and the A-StErr of the estimators always decrease when the sample size is increased.
2. The RMSE are, in general, close to the A-StErr, especially for a sample size of 5000.
3. When some bias is observed (e.g., parameters ω_2 and β_2), this bias always decreases when the sample size is increased.
4. The transition probabilities are estimated with high accuracy.

Even though the results seem satisfactory overall for the MCEM–MCML algorithm, the GARCH parameters in the second regime ω_2 and β_2 display some bias. To investigate this element, I generated 1000 trajectories with sizes of $T = 500$, 1500, 5000, 10 000 and 50 000 from a GARCH model with parameters matching those in the second regime. For each of these trajectories, the MLE of the GARCH model was estimated and the mean values of these estimates for each sample size are displayed in Table 4.

In Table 4, we can note a pattern that is analogous to that observed in Table 3 for the parameters of the second regime: the ω and β parameters are estimated with large bias for small sample sizes, but this bias is gradually decreased as the

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Table 5
Descriptive statistics.

	Mean	StDev	Skewness	Kurtosis	Minimum	Maximum
Weekly S&P500	7.2	16.6	−0.59	7.2	−16.5	10.2
Daily S&P500	−0.1	21.5	−0.12	10.5	−9.5	11.0

sample size is increased. This example demonstrates that the MLE of GARCH models can be biased in small sample sizes. It seems that the biases that are observed with regard to ω_2 and β_2 in Table 3 are normal and they are not due to a fault of the MCEM–MCML algorithm. In fact, an examination of the correlation matrix of the estimated parameters reveals that the estimates of ω_2 and β_2 are highly negatively correlated which undoubtedly complicates estimation.

4.4. Computational times

Although the MCEM–MCML algorithm is an iterative simulation-based algorithm, the estimation of a data set can be accomplished in a reasonable amount of time. For instance, it takes approximately 2, 6 and 30 min to estimate a data set with sample sizes of $T = 500$, 1500 and 5000, respectively, using simulation schedule 1 (see Table 1) on a 3.40 GHz Intel Core i7-2600 processor. The single-move Gibbs sampler implementation in C++ is efficient as it takes approximately 15, 70 and 500 s to generate 10 000 draws from the posterior distribution of the state vector at sample sizes of $T = 500$, 1500 and 5000, respectively. The maximization step of the algorithm is computationally demanding because the conditional variance at time t, $\sigma_t^2(S_{1:t})$, must be recalculated for each simulated state sequence $S_{1:t}$ and each $t = 1, \ldots, T$ whenever the parameters change. This element justifies the use of the MCML algorithm to help reduce the number of iterations when convergence of the MCEM algorithm is slow.

5. Empirical study

In this section, the MS-GARCH model is fitted to (i) weekly percentage log-returns on the S&P 500 price index from October 28, 1987 to October 31, 2012 and (ii) daily percentage log-returns on the S&P 500 price index from May 20, 1999 to April 25, 2011. Weekly data is from Wednesdays close to the following Wednesdays close to avoid most holidays and includes 1305 observations. The daily data set contains 3000 observations and allows for a comparison of the estimation results obtained with the MCEM–MCML algorithm to the Bayesian MCMC approach developed by Bauwens et al. (2011) since these authors considered the same data set. Descriptive statistics for the two financial time series are provided in Table 5 (the mean and standard deviation, abbreviated StDev, are given on an annualized basis). Similarly to Section 4, the estimation results presented in this section are reproducible with the code available on my website.

5.1. Estimation results and effectiveness of simulation schedule 2

To demonstrate the effectiveness of simulation schedule 2 (see Table 2), the parameter estimation for each data set was repeated 40 times using that strategy and compared to a brute force implementation of the MCEM–MCML algorithm (with $m_1, \ldots, m_{40} = 10\,000$, $m_{41}, \ldots, m_{50} = 50\,000$ and $m^\star = 100\,000$). The mean, median and standard deviation of the estimated parameters and log-likelihood values (denoted by log-lik) obtained over these repeated estimations are given in Table 6. The log-likelihood of the MS-GARCH model was approximated using the particle filter methodology with 100 000 particles (see Bauwens et al., 2011; Creal, 2012), which is accurate to the first decimal place.

Unconstrained estimation of MS-GARCH models with empirical data can lead to parameters being estimated on the boundary of the parameter space and result in slow convergence of the MCEM–MCML algorithm. For example, Bauwens et al. (2010) and Francq and Zakoïan (2008) fitted the MS-GARCH model to daily S&P 500 data: Bauwens et al. (2010) used the constraint $\alpha_1 = \beta_1 = 0$ in the estimation process while Francq and Zakoïan (2008) reported an estimated value of α_1 very close to zero. To obtain convergence in the interior of the parameter space, I fitted a constrained MS-GARCH model by imposing $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ in the estimation process. For weekly data, both the constrained and unconstrained versions were estimated but due to slow convergence simulation schedule 2 was not effective at estimating the unconstrained version. For daily data, the unconstrained MS-GARCH model was fitted with zero means (i.e., $\mu_1 = \mu_2 = 0$) to match the specification considered by Bauwens et al. (2011). This restriction on the means caused the MCEM–MCML algorithm to converge in the interior of the parameter space. The results of this estimation are consistent with those presented by Bauwens et al. (2011).

The results displayed in Table 6 suggest that simulation schedule 2 can be used reliably to estimate financial data when the MLE is not on the boundary of the parameter space. The estimation of the constrained model is very accurate, but more variability is observed for the unconstrained model fitted with zero means to daily data. This is due to the single-move Gibbs sampler struggling to move between highly persistent regimes. Ways to improve this sampler are discussed in the conclusion. To estimate the unconstrained model with weekly data, over 100 iterations were needed to be confident that the algorithm converged as the parameter α_2 was attracted to 0. Even though $\beta_2 > 1$, the estimated model is covariance stationary as it satisfies the conditions given by Francq et al. (2001) (see also Bauwens et al., 2010).

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Table 6
Estimation results and effectiveness of simulation schedule 2.

	μ_1	μ_2	ω_1	α_1	β_1	ω_2	α_2	β_2	p_{11}	p_{22}	log-lik
Weekly S&P 500: unconstrained model											
Brute	0.547	−2.01	0.111	0.0277	0.807	0.098	$<10^{-8}$	1.82	0.833	0.109	−2740.3
Weekly S&P 500: constrained model with $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$											
Brute	0.343	−2.80	0.0446	0.0429	0.902	2.51	0.0429	0.902	0.946	0.308	−2757.5
Median	0.346	−2.77	0.0445	0.0433	0.901	2.50	0.0433	0.901	0.945	0.308	−2757.5
Mean	0.346	−2.77	0.0446	0.0432	0.901	2.50	0.0432	0.901	0.945	0.306	−2757.5
StDev	0.007	0.12	0.0012	0.0009	0.003	0.04	0.0009	0.003	0.004	0.013	0.06
Daily S&P 500: unconstrained model with zero means											
Brute	–	–	0.0532	0.0946	0.885	0.0128	0.0194	0.953	0.999	0.999	−4476.6
Median	–	–	0.0538	0.0943	0.884	0.0125	0.0216	0.952	0.999	0.999	−4476.7
Mean	–	–	0.0609	0.0938	0.883	0.0131	0.0237	0.948	0.997	0.998	−4477.0
StDev	–	–	0.0151	0.0033	0.004	0.0035	0.0065	0.012	0.003	0.002	0.53
Daily S&P 500: constrained model with $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$											
Brute	0.0682	−1.05	0.00698	0.0337	0.942	0.527	0.0337	0.942	0.980	0.638	−4450.9
Median	0.0683	−1.06	0.00696	0.0339	0.942	0.528	0.0339	0.942	0.980	0.635	−4450.9
Mean	0.0683	−1.06	0.00695	0.0340	0.942	0.527	0.0340	0.942	0.980	0.634	−4450.9
StDev	0.0007	0.04	0.00007	0.0008	0.001	0.027	0.0008	0.001	0.000	0.025	0.05

Table 7
Weekly S&P 500: Estimated parameters and asymptotic standard errors.

	μ_1	μ_2	ω_1	ω_2	α	β	p_{11}	p_{22}	BIC
GARCH	0.209 (0.050)		0.176 (0.058)		0.131 (0.024)	0.841 (0.029)			2822.4
MS	0.281 (0.056)	−0.141 (0.167)	2.19 (0.18)	11.2 (1.0)			0.977 (0.009)	0.953 (0.017)	2815.5
Gray	0.236 (0.058)	−2.37 (0.80)	$<10^{-8}$ ($<10^{-8}$)	4.34 (2.08)	0.0698 (0.0260)	0.848 (0.040)	0.984 (0.008)	0.487 (0.122)	2805.8
MS-GARCH	0.343 (0.060)	−2.80 (0.63)	0.0446 (0.0222)	2.51 (0.52)	0.0429 (0.0205)	0.902 (0.028)	0.946 (0.022)	0.308 (0.179)	2786.2

Table 8
Daily S&P 500: Estimated parameters and asymptotic standard errors.

	μ_1	μ_2	ω_1	ω_2	α	β	p_{11}	p_{22}	BIC
GARCH	0.00223 (0.0167)		0.0125 (0.0030)		0.0759 (0.0086)	0.916 (0.009)			4510.2
MS	0.0571 (0.0190)	−0.110 (0.064)	0.631 (0.033)	4.10 (0.25)			0.989 (0.003)	0.979 (0.006)	4661.8
Gray	0.0422 (0.0164)	−1.73 (0.65)	$<10^{-8}$ ($<10^{-8}$)	0.769 (0.511)	0.0680 (0.0080)	0.908 (0.006)	0.996 (0.002)	0.507 (0.151)	4495.5
MS-GARCH	0.0682 (0.0177)	−1.05 (0.34)	0.00698 (0.00222)	0.527 (0.211)	0.0337 (0.0127)	0.942 (0.012)	0.980 (0.006)	0.638 (0.164)	4483.0

Finally, the computational time required to complete a single estimation with simulation schedule 2 on a 3.40 GHz Intel Core i7-2600 processor is approximately 35 min for the weekly data set and 110 min for the daily data set.

5.2. Comparison of fit

The fit of the constrained MS-GARCH model for the weekly and daily data sets is compared to the model of Gray (1996) and to standard MS and GARCH models in Tables 7 and 8. The MS model is a special case of the MS-GARCH model with $\alpha_1 = \alpha_2 = 0$ and $\beta_1 = \beta_2 = 0$. The Bayesian information criterion (BIC) adds a penalty of $0.5k \log T$ to the negative of the log-likelihood, where k is the number of parameters in the model and T is the number of observations. The preferred model is the one with the lowest BIC. The asymptotic standard errors of the MLE are given in parentheses.

The fit of the MS-GARCH model is superior to Gray's model according to the BIC for both data sets considered. The BIC reported for Gray's model is based on the log-likelihood of Gray's model. The log-likelihood of the MS-GARCH model

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

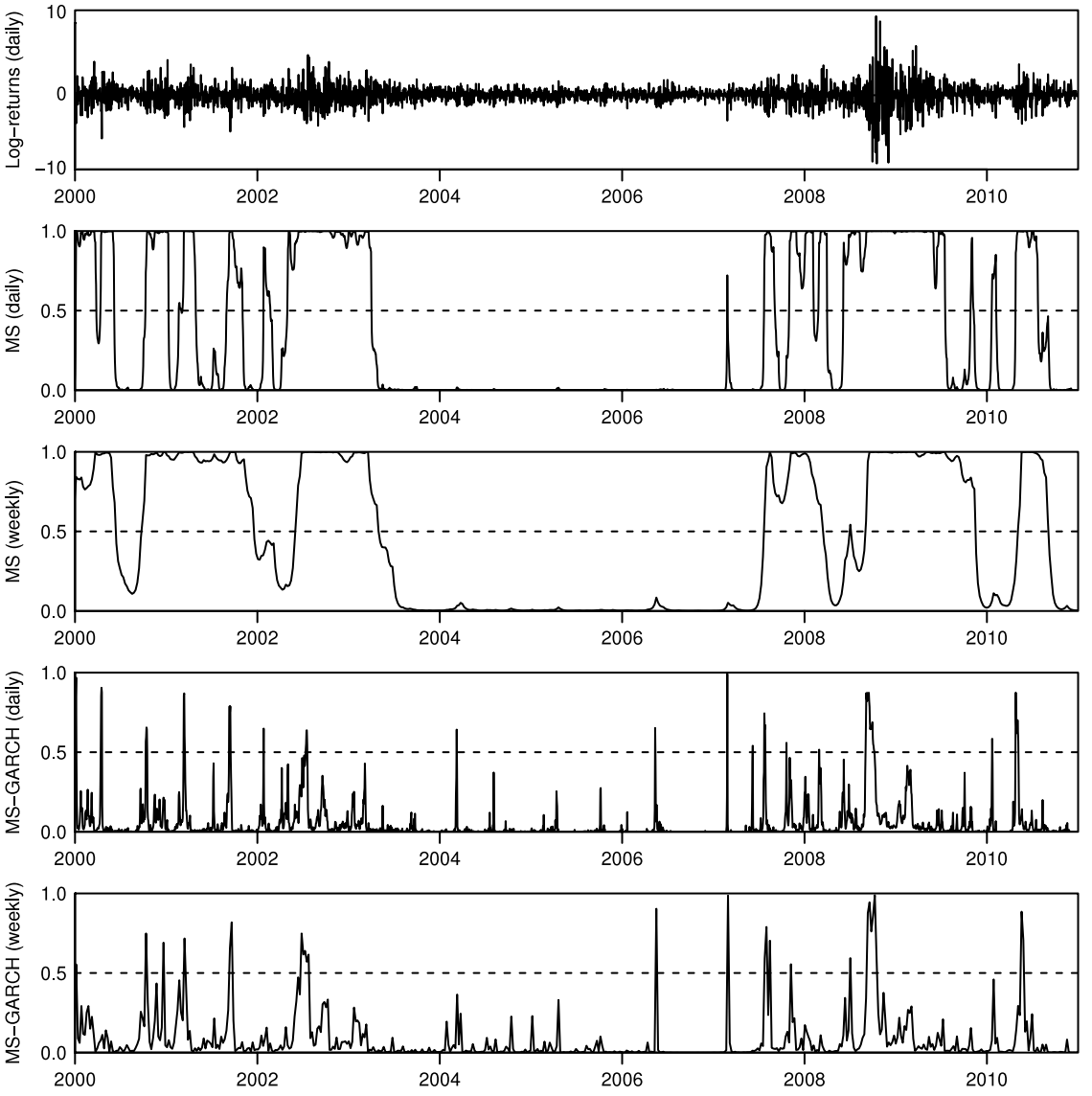

Fig. 1. S&P 500: Smoothed probabilities of being in regime two.

evaluated at the MLE of Gray's model was generally below that of the GARCH model. This implies that Gray's model can only generate a crude estimate of the MLE for the MS-GARCH model. These results are in line with those obtained in Section 4.3, where it was shown that Gray's model does not generate consistent estimates for the MS-GARCH model.

For the two time series considered, the fitted standard MS models include two persistent regimes: the first regime defines a positive mean return–low volatility state while the second regime pertains to a negative mean return–high volatility environment. When GARCH dynamics are incorporated into the MS model, the regime associated with a negative mean return becomes much less persistent (i.e., p_{22} is significantly reduced). To illustrate the difference in the role played by regime two in the MS and MS-GARCH models, we can compare the smoothed inferences of the states given by these models (see Fig. 1). For example, consider the global financial crisis that emerged in September of 2008 with the bankruptcy of Lehman Brothers and the collapse of large financial institutions around the world. From September of 2008 to August of 2009, the MS model infers that the return process is in regime two, i.e., the negative mean return–high volatility regime. During that same period, the MS-GARCH model infers that this process enters regime two at the beginning of September of 2008 and returns to regime one five weeks later. Consequently, state two in the MS-GARCH model represents a shock regime which induces a jump in the volatility process. When the model reverts back to regime one, the effect of this shock still persists in the volatility due to the GARCH dynamics. This example demonstrates that volatility persistence is captured differently in the MS and MS-GARCH models. For the MS model, it is directly tied to regime persistence, i.e., long periods of high volatility can only occur when the return process remains in regime two. For the MS-GARCH model, it is better

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

explained by the GARCH dynamics of the model since the role of the MS process is now to allow for jumps in volatility. Therefore, it is not surprising that enriching the GARCH model with a MS process offers an improved fit as the presence of these jumps is well documented in the econometric literature (e.g., Eraker et al., 2003).

6. Conclusion

A novel approach was introduced based on the MCEM and MCML algorithms to calculate the MLE of the MS-GARCH model and its effectiveness was demonstrated with a simulation study. The main contribution of this method is that it allows us to estimate the MS-GARCH model by maximum likelihood without resorting to a simplification of the model like the one used by Gray (1996). It was shown that Gray's model does not generate consistent estimates for model (1)–(3) and that the MCEM–MCML algorithm can significantly improve these estimates. The practical implementation of the algorithm was discussed and a simulation schedule was suggested to fit financial data. Finally, it was explained how to compute the variance–covariance matrix of the MLE. This paper offers many opportunities for further research.

First, the proposed algorithm can be extended to the case of asymmetric power GARCH regimes (see Ding et al., 1993) with skewed or non-normal innovations. Haas (2010) generalized the model of Haas et al. (2004) in a similar way and showed that such a specification is sometimes preferred when it is fitted to financial data. As mentioned in Section 3.5.2, automated rules for selecting the sample size at each iteration of the algorithm can be investigated. Furthermore, the single-move Gibbs sampler that was proposed to generate the states may be poorly mixing in some instances (see Frühwirth-Schnatter, 2006, Section 11.5.6). To improve this Gibbs sampler, states can be simulated in blocks instead of individually or in a random order instead of sequentially (see Levine, 2005). Simulation in blocks would increase the computational expense of the sampler by a factor of (N^b/b), where b is the block size. Consequently, it is not guaranteed that this will improve the performance of the algorithm for a fixed computational time. Recently, Bauwens et al. (2011) proposed a multi-move Gibbs sampler (also known as a forward filtering–backward sampling algorithm) for the MS-GARCH model using particle filters. This algorithm can significantly enhance the mixing properties of the single-move sampler and reduce its sensitivity to initial values.

Second, other methods to estimate the MS-GARCH model can be investigated. For example, Jacquier et al. (2007) proposed a MCMC maximum likelihood approach for latent state models that shares some similarities with the MCEM algorithm (see Doucet et al., 2002; Gaetan and Yao, 2003; Johansen et al., 2008). It involves a stochastic implementation of the M-step that can help reduce computational time and prevent convergence to a local mode of the log-likelihood. Moreover, since we can approximate the gradient and Hessian of the log-likelihood (see Section 3.4), gradient-based algorithms (see Cappé et al., 2005, Sections 10.1.3, 11.1.3 and 11.3.5) can be investigated.

Finally, the proposed algorithm applies to a univariate MS-GARCH model. The academic literature on multivariate MS-GARCH models is scarce, especially with regard to estimation. For example, Bauwens et al. (2007) and Haas et al. (2009) considered multivariate versions of the model of Haas et al. (2004). Since that model is not path dependent, estimation of these multivariate models can be done by a direct maximization of the log-likelihood. In the context of a multivariate generalization of the (path dependent) MS-GARCH model, more sophisticated methods are needed. It would thus be interesting to extend the proposed algorithm to a multivariate setting or to perform estimation using a Bayesian approach.

Acknowledgments

I thank the Associate Editor, two referees, Mathieu Boudreault and Manuel Morales for their constructive comments which helped me improve this paper. I am also grateful to Brian Hartman and Eden Tsang who proofread this article and Arnaud Dufays for providing me with the data set used by Bauwens et al. (2011). Finally, I would like to acknowledge the financial support provided by the Natural Sciences and Engineering Research Council of Canada and the Society of Actuaries.

Appendix A. Validity of the EM algorithm for the MS-GARCH model

Strictly speaking, the MLE in mixture or MS models may not exist because of an unbounded likelihood. For example, consider a mixture of two normal distributions in which one of the normal densities has a mean exactly equal to one of the observations with a variance in the vicinity of 0. In this case, the likelihood is unbounded on the boundary of the parameter space so that a global maximum of the likelihood does not exist (see Frühwirth-Schnatter, 2006, Sections 6.1.2 and 6.1.3). However, from a practical perspective, the area of the parameter space which creates this anomaly is of no interest since it occurs when one of the states is essentially trying to give discrete mass to one of the observations. Therefore, we may constrain the parameter space to exclude these pathological cases by, for example, posing a strictly positive lower bound on the variances. With this restriction, Kiefer (1978) proves that there exists a bounded local maximizer of the likelihood which is consistent, efficient and asymptotically normal (see also Hathaway, 1985; Ingrassia and Rocci, 2007).

The convergence of the EM algorithm is discussed in great detail by Wu (1983) who proves that if $Q(\theta \mid \theta')$ is continuous in both θ and θ' then (under regularity conditions) all limit points of the EM algorithm are stationary points of the likelihood.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

The following regularity conditions are needed (Ω denotes the parameter space):

$$\Omega \text{ is a subset in the } d\text{-dimensional Euclidean space } \mathbb{R}^d,$$
$$\Omega_{\theta_0} = \{\theta \in \Omega : f(y \mid \theta) \geq f(y \mid \theta_0)\} \text{ is compact for any } f(y \mid \theta_0) > -\infty, \tag{A.1}$$
$$\text{each parameter iterate } \theta^{(r)} \text{ is in the interior of } \Omega,$$
$$f(y \mid \theta) \text{ is continuous in } \Omega \text{ and differentiable in the interior of } \Omega. \tag{A.2}$$

A consequence of these conditions is that any sequence of likelihood values generated by the EM algorithm is bounded above. The compactness assumption (A.1) is difficult to verify but if we constrain the parameter space to exclude cases related to an unbounded likelihood, this assumption should hold for the MS-GARCH model (we can require that the GARCH parameters in all of the states be greater than some lower bound strictly greater than 0). Moreover, condition (A.2) is satisfied for the MS-GARCH model since (see Eq. (4))

$$f(y \mid \theta) = \sum_{S} \prod_{t=1}^{T} \left[\sigma_{t}^{-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_t - \mu_{S_t})^2}{2\sigma_t^2} \right) p_{S_{t-1},S_t} \right],$$

is continuous in Ω and differentiable in the interior of Ω. Finally,

$$\begin{aligned} Q(\theta \mid \theta') &= \sum_{S} \log[f(y, S \mid \theta)] p(S \mid y, \theta') \\ &= \frac{1}{f(y \mid \theta')} \sum_{S} \log[f(y, S \mid \theta)] f(y, S \mid \theta'), \end{aligned}$$

is continuous in both θ and θ' since $f(y, S \mid \theta)$ and $f(y \mid \theta)$ are positive continuous functions of θ.

This implies that it is valid to use the EM algorithm to estimate the MS-GARCH model since all limit points generated by the algorithm are stationary points of the likelihood. There is no guarantee that the stationary point is a local maximum, but convergence to a local minimum or a saddle point only occurs for some pathological cases so that in almost all instances, the stationary point is a local maximum (see McLachlan and Krishnan, 2008, Section 3.6). If the likelihood function is unimodal in θ and there is only one stationary point then the EM algorithm will converge to the global (unique) maximizer of the likelihood. However, the likelihood in mixture and MS models can have many modes and convergence of the EM sequence to a particular mode may depend on the choice of starting values. Hence, it is best to start the EM algorithm with different sets of starting values to avoid a suboptimal local maximum.

Appendix B. Analytical maximization of term 2 in expression (10)

Maximization of term 2 in expression (10) with respect to the transition probability parameters $\{p_{jk}\}_{j,k=1}^{N}$ subject to the constraint $\sum_{k=1}^{N} p_{jk} = 1, j = 1, \ldots, N$ can be done analytically. Let $f_{jk}^{(i)}$ denote the number of transitions from state j to state k in the ith simulated state vector $S^{(i)}$ and let $f_{jk} = \sum_{i=1}^{m_r} f_{jk}^{(i)}$ (i.e., f_{jk} is the total number of transitions from state j to state k in all of the m_r simulated state vectors). Then, we may rewrite term 2 in expression (10) as

$$\begin{aligned} \text{term 2} &= \frac{1}{m_r} \sum_{t=1}^{T} \sum_{i=1}^{m_r} \log(p_{S_{t-1}^{(i)},S_t^{(i)}}) \\ &= \frac{1}{m_r} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{i=1}^{m_r} f_{jk}^{(i)} \log(p_{jk}) \\ &= \frac{1}{m_r} \sum_{j=1}^{N} \sum_{k=1}^{N} f_{jk} \log(p_{jk}) \\ &= \frac{1}{m_r} \sum_{j=1}^{N} \left[\sum_{k=1}^{N-1} f_{jk} \log(p_{jk}) + f_{jN} \log\left(1 - \sum_{k=1}^{N-1} p_{jk} \right) \right]. \end{aligned} \tag{B.1}$$

For each j, the expression inside the summation of Eq. (B.1) only involves transition probabilities of the jth row of the transition matrix. Therefore, we may find the optimal values of the transition probabilities for each row independently. This implies that for row j we must find the values of $\{p_{jk}\}_{k=1}^{N-1}$ which maximize the expression

$$\sum_{k=1}^{N-1} f_{jk} \log(p_{jk}) + f_{jN} \log\left(1 - \sum_{k=1}^{N-1} p_{jk} \right).$$

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

This can be done by using straightforward calculus steps and the closed-form expressions for the maximizers of the transition probabilities are

$$p_{jk} = \frac{f_{jk}}{\sum_{l=1}^{N} f_{jl}}, \quad j,k = 1,\ldots,N.$$

Appendix C. Technical details related to the MCML-step of Algorithm 1

To obtain $\hat{\theta}$, we must calculate $w^{(i)}_{\theta|\theta^\star}$, $i = 1,\ldots,m^\star$, which are defined in Eq. (6). From a numerical perspective, it is best to calculate $\log(w^{(i)}_{\theta|\theta^\star})$, where

$$\begin{aligned}\log(w^{(i)}_{\theta|\theta^\star}) &= \log f(y, S^{(i)} \mid \theta) - \log f(y, S^{(i)} \mid \theta^\star) \\ &= \frac{1}{2}\sum_{t=1}^{T}\left[\log(\sigma_t^{\star(i)})^2 - \log(\sigma_t^{(i)})^2 + \frac{(y_t - \mu^\star_{S_t^{(i)}})^2}{(\sigma_t^{\star(i)})^2} - \frac{(y_t - \mu_{S_t^{(i)}})^2}{(\sigma_t^{(i)})^2}\right] + \sum_{j=1}^{N}\sum_{k=1}^{N} f^{(i)}_{jk}\log(p_{jk}/p^\star_{jk}). \end{aligned} \tag{C.1}$$

The starred quantities in expression (C.1) must be calculated based on $\theta^\star$ and these that are not starred are based on θ. As in Appendix B, $f^{(i)}_{jk}$ denotes the number of transitions from state j to state k in the ith simulated state vector $S^{(i)}$.

To improve the performance of the numerical optimization in expression (11), the gradient of $\log\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}$ with respect to θ can be calculated in closed form using the following relation:

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\right] &= \sum_{i=1}^{m^\star}\frac{\partial f(y,S^{(i)}\mid\theta)/\partial\theta}{f(y,S^{(i)}\mid\theta^\star)} = \sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\log f(y,S^{(i)}\mid\theta) \\ &= \sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\log f(y\mid S^{(i)},\theta) + \sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\log p(S^{(i)}\mid\theta).\end{aligned}$$

Letting $\bar{w}^{(i)}_{\theta|\theta^\star} = w^{(i)}_{\theta|\theta^\star}/\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}$, we obtain

$$\begin{aligned}\frac{\partial}{\partial\theta}\left[\log\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}\right] &= \sum_{i=1}^{m^\star}\bar{w}^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\log f(y\mid S^{(i)},\theta) + \sum_{i=1}^{m^\star}\bar{w}^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\log p(S^{(i)}\mid\theta) \\ &= \text{gradient 1} + \text{gradient 2}.\end{aligned} \tag{C.2}$$

Gradients 1 and 2 simplify to

$$\text{gradient 1} = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{m^\star}\bar{w}^{(i)}_{\theta|\theta^\star}\frac{\partial}{\partial\theta}\left[\log(\sigma_t^{(i)})^2 + \frac{(y_t - \mu_{S_t^{(i)}})^2}{(\sigma_t^{(i)})^2}\right], \tag{C.3}$$

$$\text{gradient 2} = \sum_{j=1}^{N}\sum_{k=1}^{N}\left[\left(\sum_{i=1}^{m^\star}\bar{w}^{(i)}_{\theta|\theta^\star} f^{(i)}_{jk}\right)\frac{\partial}{\partial\theta}\log(p_{jk})\right]. \tag{C.4}$$

Gradient 1 can be calculated recursively without difficulty while gradient 2 can be computed directly. Note that the gradient of $\log\sum_{i=1}^{m^\star} w^{(i)}_{\theta|\theta^\star}$ is approximately equal to the gradient of the log-likelihood (see Eq. (7)). It is also closely related to the gradient of $\widehat{Q}(\theta\mid\theta^\star)$ which can be obtained from expressions (C.3) and (C.4) by replacing $\bar{w}^{(i)}_{\theta|\theta^\star}$, $i = 1,\ldots,m^\star$, with $1/m^\star$. In this case, gradient 1 is the gradient of term 1 in expression (10).

References

Bauwens, L., Dufays, A., Rombouts, J.V., 2011. Marginal likelihood for Markov-switching and change-point GARCH models. CORE Discussion Papers 2011013. Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
Bauwens, L., Hafner, C.M., Rombouts, J.V.K., 2007. Multivariate mixed normal conditional heteroskedasticity. Computational Statistics & Data Analysis 51, 3551–3566.
Bauwens, L., Preminger, A., Rombouts, J.V.K., 2010. Theory and inference for a Markov switching GARCH model. The Econometrics Journal 13, 218–244.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.
Caffo, B.S., Jank, W., Jones, G.L., 2005. Ascent-based Monte Carlo expectation–maximization. Journal of the Royal Statistical Society. Series B. Statistical Methodology 67, 235–251.
Cappé, O., Douc, R., Moulines, E., Robert, C., 2002. On the convergence of the Monte Carlo maximum likelihood method for latent variable models. Scandinavian Journal of Statistics. Theory and Applications 29, 615–635.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026

Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. In: Springer Series in Statistics, Springer, New York.
Cont, R., 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1, 223–236.
Creal, D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econometric Reviews 31, 245–296.
Danielsson, J., Richard, J.F., 1993. Accelerated Gaussian importance sampler with application to dynamic latent variable models. Journal of Applied Econometrics 8, S153–S173.
Ding, Z., Granger, C.W.J., Engle, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.
Doucet, A., Godsill, S.J., Robert, C.P., 2002. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. Statistics and Computing 12, 77–84.
Dueker, M.J., 1997. Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business & Economic Statistics 15, 26–34.
Eddelbuettel, D., François, R., 2011. Rcpp: Seamless R and C++ integration. Journal of Statistical Software 40, 1–18.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.
Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. The Journal of Finance 58, 1269–1300.
Francq, C., Roussignol, M., Zakoïan, J.M., 2001. Conditional heteroskedasticity driven by hidden Markov chains. Journal of Time Series Analysis 22, 197–220.
Francq, C., Zakoïan, J.M., 2005. The L^2-structures of standard and switching-regime GARCH models. Stochastic Processes and their Applications 115, 1557–1582.
Francq, C., Zakoïan, J.M., 2008. Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference. Computational Statistics & Data Analysis 52, 3027–3046.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. In: Springer Series in Statistics, Springer, New York.
Gaetan, C., Yao, J.F., 2003. A multiple-imputation Metropolis version of the EM algorithm. Biometrika 90, 643–654.
Geyer, C.J., 1994. On the convergence of Monte Carlo maximum likelihood calculations. Journal of the Royal Statistical Society. Series B. Methodological 56, 261–274.
Geyer, C.J., 1996. Estimation and optimization of functions. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 241–258.
Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.
Haas, M., 2010. Skew-normal mixture and Markov-switching GARCH processes. Studies in Nonlinear Dynamics & Econometrics 14, Article 1.
Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to Markov-switching GARCH models. Journal of Financial Econometrics 2, 493–530.
Haas, M., Mittnik, S., Paolella, M.S., 2009. Asymmetric multivariate normal mixture GARCH. Computational Statistics & Data Analysis 53, 2129–2154.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Hamilton, J.D., 2008. Regime switching models. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke.
Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.
Hathaway, R.J., 1985. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. The Annals of Statistics 13, 795–800.
Ingrassia, S., Rocci, R., 2007. Constrained monotone EM algorithms for finite mixture of multivariate Gaussians. Computational Statistics & Data Analysis 51, 5339–5351.
Jacquier, E., Johannes, M., Polson, N., 2007. MCMC maximum likelihood for latent state models. Journal of Econometrics 137, 615–640.
Johansen, A.M., Doucet, A., Davy, M., 2008. Particle methods for maximum likelihood estimation in latent variable models. Statistics and Computing 18, 47–57.
Jones, G.L., Hobert, J.P., 2001. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. Statistical Science 16, 312–334.
Kiefer, N.M., 1978. Discrete parameter variation: efficient estimation of a switching regression model. Econometrica 46, 427–434.
Klaassen, F., 2002. Improving GARCH volatility forecasts with regime-switching GARCH. Empirical Economics 27, 363–394.
Lamoureux, C.G., Lastrapes, W.D., 1990. Persistence in variance, structural change, and the GARCH model. Journal of Business & Economic Statistics 8, 225–234.
Levine, R.A., 2005. A note on Markov chain Monte Carlo sweep strategies. Journal of Statistical Computation and Simulation 75, 253–262.
Levine, R.A., Fan, J., 2004. An automated (Markov chain) Monte Carlo EM algorithm. Journal of Statistical Computation and Simulation 74, 349–359.
Louis, T.A., 1982. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. Series B. Methodological 44, 226–233.
McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association 92, 162–170.
McLachlan, G.J., Krishnan, T., 2008. The EM Algorithm and Extensions, second ed. In: Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ.
Mikosch, T., Starica, C., 2004. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. The Review of Economics and Statistics 86, 378–390.
Pitt, M.K., 2002. Smooth particle filters for likelihood evaluation and maximisation. Warwick Economic Research Papers 651. Department of Economics, University of Warwick.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.
Wei, G.C.G., Tanner, M.A., 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. Journal of the American Statistical Association 85, 699–704.
Wu, C.F.J., 1983. On the convergence properties of the EM algorithm. The Annals of Statistics 11, 95–103.
Yao, W., 2011. Label switching and its simple solutions for frequentist mixture models. Manuscript (submitted for publication). Available at http://www-personal.ksu.edu/~wxyao/research.html.
Zucchini, W., MacDonald, I.L., 2009. Hidden Markov Models for Time Series: An Introduction Using R. In: Monographs on Statistics and Applied Probability, vol. 110. Chapman & Hall/CRC, Boca Raton, FL.

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026